Table of Contents

# Relazione tra Sentiment nei Commenti e Veridicità delle Notizie

# Analisi del Dataset PHEME



Copertina del progetto

**Università:** Corso di Digital Humanities  
**Data:** 3 maggio 2025  
**Autore:** Simone

## Indice Generale

1. [Introduzione](01_introduzione.md)
   * Contesto e motivazione
   * Il problema della disinformazione online
   * Obiettivi dello studio
   * Ipotesi di ricerca
2. [Dataset e Metodologia](02_dataset_metodologia.md)
   * Dataset PHEME
   * Preprocessing dei dati
   * Estrazione delle feature
   * Metodi di analisi
3. [Analisi Esplorativa](03_analisi_esplorativa.md)
   * Statistiche descrittive
   * Distribuzione delle feature
   * Pattern principali
4. [Analisi Statistica](04_analisi_statistica.md)
   * Test di ipotesi
   * Correlazioni
   * Significatività e effect size
5. [Modelli Predittivi](05_modelli_predittivi.md)
   * Regressione logistica
   * Random Forest
   * Confronto tra modelli
   * Importanza delle feature
6. [Risultati e Discussione](06_risultati_discussione.md)
   * Sintesi dei risultati principali
   * Interpretazione nel contesto della ricerca
   * Rilevanza pratica e teorica
7. [Limitazioni e Validazione](07_limitazioni_validazione.md)
   * Limitazioni metodologiche
   * Validazione dei risultati
   * Generalizzabilità
8. [Conclusioni](08_conclusioni.md)
   * Sintesi finale
   * Implicazioni
   * Direzioni future
9. [Bibliografia](09_bibliografia.md)

## Sommario Esecutivo

Questo studio esplora la relazione tra i pattern di sentiment espressi nei commenti alle notizie e la veridicità dell’informazione originale, utilizzando il dataset PHEME di conversazioni Twitter. Attraverso un approccio multi-metodologico che combina analisi statistica e machine learning, abbiamo indagato se esistano differenze sistematiche nelle reazioni linguistiche e affettive alle notizie vere rispetto a quelle false.

I risultati mostrano differenze statisticamente significative ma con effect size trascurabile nelle feature di sentiment tra i commenti a notizie vere e false. I modelli non lineari (Random Forest) catturano relazioni più complesse rispetto ai modelli lineari (Regressione Logistica), con un miglioramento dell’AUC da 0.54 a 0.93. Le feature di leggibilità e acculturazione, in particolare il culture\_score, si rivelano più predittive rispetto alle pure feature di sentiment.

Questi risultati suggeriscono che l’analisi del sentiment da sola è insufficiente per identificare efficacemente le fake news, evidenziando la necessità di approcci multidimensionali che considerino la complessità linguistica e le dinamiche conversazionali. Lo studio contribuisce alla comprensione dei meccanismi di diffusione della disinformazione online e alle strategie per contrastarla.

# Sommario Esecutivo

## Studio sulla Relazione tra Sentiment nei Commenti e Veridicità delle Notizie

### Panoramica

Questo studio ha esplorato la relazione tra i pattern linguistici presenti nei commenti sui social media e la veridicità delle notizie originali, utilizzando il dataset PHEME di conversazioni Twitter. L’obiettivo era determinare se esistano differenze sistematiche nelle reazioni linguistiche e affettive alle notizie vere rispetto a quelle false, e se queste differenze possano essere sfruttate per identificare la disinformazione.

### Metodologia

La ricerca ha adottato un approccio multi-metodologico:

1. **Estrazione di feature linguistiche** dai commenti:
   * Feature di sentiment (polarità, soggettività)
   * Feature di stance (atteggiamento verso il contenuto originale)
   * Feature di leggibilità e acculturazione (tra cui il culture\_score)
2. **Analisi statistica** per identificare differenze significative tra i gruppi
3. **Modellazione predittiva** con approcci lineari e non lineari:
   * Regressione logistica (modello lineare)
   * Random Forest (modello non lineare)
4. **Confronto tra set di feature** per valutare il contributo di diverse categorie linguistiche

### Principali Risultati

1. **Differenze statisticamente significative ma con effect size limitato**
   * Differenze significative in polarità del sentiment, soggettività, stance e culture\_score
   * Tutti gli effect size trascurabili (<0.1), indicando limitata rilevanza pratica di singole feature
2. **Superiorità marcata dei modelli non lineari**
   * Random Forest: AUC 0.932
   * Regressione logistica: AUC 0.542
   * Incremento: +0.390, suggerendo relazioni prevalentemente non lineari
3. **Maggiore rilevanza delle feature di leggibilità**
   * Set di feature di leggibilità (AUC 0.571) superiore a sentiment (AUC 0.559)
   * Culture\_score emerso come la feature linguistica più importante
4. **Rischio di overfitting**
   * Alta importanza degli identificatori nei modelli suggerisce potenziale specificità al dataset
   * Calo significativo di performance escludendo ID (da AUC 0.932 a 0.682)
5. **Valore dell’integrazione di diverse dimensioni**
   * Performance ottimale con combinazione di feature di sentiment, stance e leggibilità
   * Approccio multidimensionale essenziale per catturare la complessità del fenomeno

### Implicazioni

1. **Per la ricerca sulla disinformazione**
   * Necessità di approcci non lineari per catturare relazioni complesse
   * Importanza della dimensione cognitiva (complessità linguistica, acculturazione) oltre alla dimensione emotiva
   * Valore di prospettive multidimensionali che integrino diverse caratteristiche linguistiche
2. **Per sistemi di fact-checking**
   * Integrare indicatori di complessità linguistica e acculturazione oltre all’analisi del sentiment
   * Utilizzare modelli non lineari capaci di catturare pattern complessi
   * Considerare il contesto conversazionale e posizionale dei commenti
3. **Per l’educazione ai media**
   * Enfatizzare lo sviluppo del pensiero critico e della complessità argomentativa
   * Sensibilizzare alla qualità del linguaggio come possibile indicatore di affidabilità
   * Educare sul ruolo delle emozioni nella diffusione della disinformazione

### Limitazioni

1. **Dataset sbilanciato** (93% notizie vere vs 7% false)
2. **Rischio di overfitting** sui dati specifici
3. **Analisi prevalentemente statica** che non considera pienamente dinamiche temporali
4. **Specificità contestuale** a Twitter e agli eventi coperti
5. **Limitazioni nelle tecniche di analisi del sentiment**

### Conclusione

Questo studio dimostra che esistono differenze sistematiche nei pattern linguistici delle reazioni a notizie vere e false, ma queste differenze sono meglio catturate da modelli non lineari e approcci multidimensionali. Il livello di acculturazione e complessità linguistica nei commenti emerge come un indicatore potenzialmente più informativo della veridicità rispetto alle pure reazioni emotive.

I risultati suggeriscono che l’analisi dei commenti può fornire segnali diagnostici utili per l’identificazione della disinformazione, ma richiede approcci sofisticati che considerino simultaneamente diverse dimensioni linguistiche e le loro complesse interazioni. Questo apre nuove prospettive sia per la ricerca sulla disinformazione sia per lo sviluppo di strumenti pratici di fact-checking basati sull’analisi delle conversazioni online.

# 1. Introduzione

## Contesto e Motivazione

Nell’era della comunicazione digitale, la disinformazione rappresenta una delle sfide più significative per la società contemporanea. La rapidità e l’ampiezza con cui le informazioni false possono diffondersi attraverso le piattaforme social hanno implicazioni profonde sul dibattito pubblico, sui processi democratici e sulla coesione sociale. Comprendere i meccanismi che facilitano o ostacolano la propagazione della disinformazione è dunque diventato un obiettivo di ricerca prioritario in diverse discipline, dalle scienze dell’informazione alla psicologia sociale, dalle scienze politiche all’informatica.

Il presente studio si inserisce in questo filone di ricerca, adottando un approccio innovativo che sposta il focus dalle caratteristiche intrinseche delle fake news ai pattern di risposta che queste generano negli utenti. L’ipotesi di partenza è che le notizie false possano suscitare reazioni linguistiche e affettive diverse rispetto alle notizie vere, e che queste differenze possano essere rilevate attraverso l’analisi computazionale del linguaggio.

## Il Problema della Disinformazione Online

La disinformazione online si è evoluta significativamente negli ultimi anni, diventando sempre più sofisticata nei contenuti e nelle strategie di diffusione. Come evidenziato da Vosoughi et al. (2018) in uno studio pubblicato su Science, le notizie false si diffondono più velocemente, più ampiamente e più in profondità rispetto alle notizie vere, grazie a meccanismi di engagement che sfruttano bias cognitivi e trigger emotivi.

Tradizionalmente, l’identificazione delle fake news si è basata su:

1. **Approcci basati sul contenuto**: analisi del testo della notizia, del suo stile linguistico, della sua struttura
2. **Approcci basati sulla fonte**: valutazione dell’affidabilità della fonte, storia di pubblicazioni precedenti
3. **Approcci basati sulla diffusione**: pattern di propagazione attraverso le reti sociali

Tuttavia, questi approcci presentano limitazioni significative. I metodi basati sul contenuto possono essere aggirabili attraverso tecniche di scrittura sempre più sofisticate; quelli basati sulla fonte possono fallire di fronte a nuovi siti o account creati ad hoc; quelli basati sulla diffusione richiedono dati longitudinali non sempre disponibili.

### Gap nella Letteratura

Una dimensione meno esplorata, ma potenzialmente rivelatrice, riguarda i pattern di risposta che le notizie false generano negli utenti dei social media. Mentre diversi studi hanno analizzato come le fake news si diffondono attraverso le reti sociali (Vosoughi et al., 2018), meno attenzione è stata dedicata alle caratteristiche linguistiche delle reazioni degli utenti.

Gonzalez-Bailon et al. (2021) hanno suggerito che le reazioni emotive alle fake news tendono ad essere più intense e polarizzate, ma queste osservazioni non sono state sistematicamente testate su dataset diversificati. Inoltre, pochi studi hanno analizzato come la complessità linguistica e il livello di acculturazione nei commenti possano correlarsi con la veridicità dell’informazione originale.

Zubiaga et al. (2016), nello studio che ha portato alla creazione del dataset PHEME, hanno analizzato le conversazioni su Twitter relative a diversi eventi di attualità, focalizzandosi però principalmente sui pattern di diffusione piuttosto che sulle caratteristiche linguistiche delle reazioni. Il nostro studio intende colmare questa lacuna, esplorando sistematicamente come il sentiment, la stance e le caratteristiche di leggibilità dei commenti possano differire tra notizie vere e false.

## Obiettivi dello Studio

Il presente studio si propone di:

1. **Verificare l’esistenza di differenze significative** nei pattern di sentiment tra i commenti alle notizie vere e quelli alle notizie false
2. **Quantificare la forza delle associazioni** tra caratteristiche linguistiche dei commenti e veridicità delle notizie
3. **Confrontare modelli predittivi lineari e non lineari** per determinare la natura delle relazioni tra feature linguistiche e veridicità
4. **Identificare le feature linguistiche più rilevanti** per distinguere tra reazioni a notizie vere e false
5. **Valutare il potenziale predittivo** di diversi set di feature, con particolare attenzione al confronto tra feature di sentiment e di leggibilità/acculturazione

Questi obiettivi non sono solo di interesse teorico, ma hanno anche importanti implicazioni pratiche. Comprendere come gli utenti reagiscono linguisticamente a notizie vere e false potrebbe contribuire a:

* Sviluppare sistemi più accurati per l’identificazione automatica delle fake news
* Creare strumenti di fact-checking che considerino non solo il contenuto originale ma anche le reazioni che genera
* Progettare interventi educativi mirati per aumentare la resilienza degli utenti alla disinformazione
* Approfondire la conoscenza dei meccanismi cognitivi e sociali coinvolti nella diffusione della disinformazione

## Ipotesi di Ricerca

Il nostro studio è guidato da cinque ipotesi principali:

1. **Ipotesi 1**: Esistono differenze statisticamente significative nel sentiment (polarità e soggettività) dei commenti alle notizie vere rispetto a quelli alle notizie false.
   * **H0**: Non c’è differenza significativa nel sentiment tra i due gruppi
   * **H1**: Il sentiment differisce significativamente tra i due gruppi
2. **Ipotesi 2**: Esistono differenze statisticamente significative nella stance (atteggiamento) dei commenti rispetto alla notizia originale tra thread di notizie vere e false.
   * **H0**: Non c’è differenza significativa nella stance tra i due gruppi
   * **H1**: La stance differisce significativamente tra i due gruppi
3. **Ipotesi 3**: Esistono differenze statisticamente significative nelle misure di leggibilità e acculturazione (in particolare nel culture\_score) dei commenti tra notizie vere e false.
   * **H0**: Non c’è differenza significativa nelle misure di leggibilità tra i due gruppi
   * **H1**: Le misure di leggibilità differiscono significativamente tra i due gruppi
4. **Ipotesi 4**: Le feature di leggibilità e acculturazione hanno un maggior potere predittivo sulla veridicità rispetto alle pure feature di sentiment.
   * **H0**: Le feature di leggibilità non sono più predittive delle feature di sentiment
   * **H1**: Le feature di leggibilità sono più predittive delle feature di sentiment
5. **Ipotesi 5**: Modelli non lineari catturano relazioni significativamente più forti tra feature linguistiche e veridicità rispetto ai modelli lineari.
   * **H0**: Non c’è differenza significativa nella performance tra modelli lineari e non lineari
   * **H1**: I modelli non lineari hanno una performance significativamente migliore

Queste ipotesi sono state formulate sulla base di studi precedenti che suggeriscono che le reazioni alle fake news possono essere caratterizzate da maggiore emotività, minor complessità linguistica e minor riflessione critica (Pennycook & Rand, 2019). Tuttavia, queste osservazioni non sono state sistematicamente testate in contesti di social media e conversazioni online.

Nel capitolo successivo, descriveremo in dettaglio il dataset utilizzato e la metodologia adottata per testare queste ipotesi.

# 2. Dataset e Metodologia

## Dataset PHEME

### Origine e Struttura

Il PHEME dataset (Zubiaga et al., 2016) è una risorsa sviluppata specificamente per lo studio della diffusione di rumour sui social media. Contiene conversazioni Twitter relative a diversi eventi di attualità, con annotazioni sulla veridicità delle affermazioni e sulla struttura conversazionale dei thread.

Il dataset prende il nome dal personaggio mitologico greco Pheme, personificazione delle voci e dei rumour, una scelta simbolica che riflette l’obiettivo del progetto di studiare la diffusione di informazioni non verificate online.

**Caratteristiche principali**: - 6.425 thread di conversazione - 105.354 tweet totali - Copertura di eventi diversi (Charlie Hebdo, Ferguson, Germanwings crash, ecc.) - Annotazioni manuali di veridicità (true, false, unverified) - Metadati conversazionali (relazioni di risposta, posizione nel thread)

### Distribuzione del Target

Una caratteristica importante del dataset è lo sbilanciamento significativo tra le classi di veridicità:

Vere: 93% (5.973 thread)  
False: 7% (452 thread)

Questo sbilanciamento riflette la realtà dei social media, dove le notizie verificate tendono ad essere più numerose, ma rappresenta anche una sfida metodologica che richiede tecniche specifiche per garantire risultati affidabili.

### Eventi Coperti

Il dataset include tweet relativi a cinque eventi principali:

1. **Charlie Hebdo**: l’attacco terroristico alla sede del settimanale satirico francese nel gennaio 2015
2. **Sydney Siege**: la crisi degli ostaggi a Sydney nel dicembre 2014
3. **Ferguson**: le proteste seguite all’uccisione di Michael Brown in Missouri nel 2014
4. **Ottawa Shooting**: la sparatoria al Parlamento canadese nell’ottobre 2014
5. **Germanwings Crash**: l’incidente aereo sulle Alpi francesi nel marzo 2015

Questa diversità di eventi permette di analizzare le reazioni a diverse tipologie di notizie (terrorismo, proteste sociali, incidenti) in contesti culturali e geografici diversi.

### Caratteristiche dei Thread Conversazionali

I thread nel dataset PHEME sono strutturati gerarchicamente: - **Tweet sorgente**: il post originale che contiene la notizia - **Reazioni dirette**: risposte immediate al tweet sorgente - **Reazioni indirette**: risposte alle risposte, che formano conversazioni ramificate

Questa struttura conversazionale è stata preservata nell’analisi attraverso feature come reaction\_index, che indica la posizione del commento nella catena di risposte.

## Metodologia

### Approccio Generale

La nostra metodologia segue un approccio rigoroso basato sui principi del metodo scientifico:

1. **Formulazione delle ipotesi**: definizione chiara delle domande di ricerca e delle ipotesi da testare
2. **Acquisizione e preparazione dei dati**: download, pulizia e strutturazione del dataset PHEME
3. **Estrazione delle feature**: calcolo di metriche linguistiche, di sentiment e di leggibilità
4. **Analisi esplorativa**: esame delle distribuzioni e identificazione di pattern preliminari
5. **Test statistici**: verifica formale delle ipotesi tramite test appropriati
6. **Modellazione predittiva**: sviluppo e confronto di modelli lineari e non lineari
7. **Interpretazione**: analisi critica dei risultati nel contesto della letteratura esistente
8. **Validazione**: verifica della robustezza dei risultati attraverso tecniche di cross-validation

Tutte le analisi sono state condotte utilizzando Python 3.10 e librerie specializzate per l’analisi dei dati e il natural language processing (pandas, scikit-learn, nltk, TextBlob, textstat).

### Preprocessing dei Dati

La preparazione dei dati è stata una fase critica per garantire risultati affidabili. Il processo ha incluso:

#### 1. Pulizia dei testi

def clean\_text(text):  
 """Pulisce e normalizza il testo del tweet."""  
 if not text or pd.isna(text):  
 return ""  
   
 # Rimozione URL  
 text = re.sub(r'https?://\S+|www\.\S+', '', text)  
   
 # Rimozione menzioni e hashtag  
 text = re.sub(r'@\w+', '', text)  
 text = re.sub(r'#(\w+)', r'\1', text)  
   
 # Normalizzazione  
 text = text.lower()  
   
 # Rimozione caratteri speciali  
 text = re.sub(r'[^\w\s]', '', text)  
   
 # Rimozione spazi multipli  
 text = re.sub(r'\s+', ' ', text).strip()  
   
 return text

Le operazioni di pulizia hanno incluso: - Rimozione di URL e link tramite espressioni regolari - Rimozione di menzioni (@username) e hashtag - Normalizzazione di emoji e caratteri speciali - Correzione di errori comuni di ortografia

#### 2. Normalizzazione linguistica

def normalize\_text(text, remove\_stopwords=True):  
 """Normalizza il testo: tokenizzazione, rimozione stopwords, lemmatizzazione."""  
 if not text or len(text) < 3:  
 return ""  
   
 # Tokenizzazione  
 tokens = word\_tokenize(text)  
   
 # Rimozione stopwords  
 if remove\_stopwords:  
 stop\_words = set(stopwords.words('english'))  
 tokens = [token for token in tokens if token not in stop\_words]  
   
 # Lemmatizzazione  
 lemmatizer = WordNetLemmatizer()  
 tokens = [lemmatizer.lemmatize(token) for token in tokens]  
   
 return ' '.join(tokens)

Il processo di normalizzazione ha compreso: - Conversione a minuscolo - Rimozione di stop words - Lemmatizzazione per ridurre le parole alla forma base - Tokenizzazione per analisi a livello di parola

#### 3. Gestione dei valori mancanti

Abbiamo adottato criteri rigorosi per la gestione dei dati incompleti o inadeguati: - Esclusione di tweet con testi insufficienti (< 10 caratteri) - Esclusione di thread con meno di 3 reazioni - Documentazione di tutti i criteri di esclusione per trasparenza

#### 4. Strutturazione gerarchica

Per preservare il contesto conversazionale, abbiamo: - Ricostruito la struttura dei thread - Assegnato indici posizionali a ciascun tweet - Associato ciascun commento al proprio tweet sorgente

### Estrazione delle Feature

Per catturare diverse dimensioni delle reazioni linguistiche, abbiamo estratto tre categorie principali di feature:

#### 1. Feature di Sentiment Analysis

Utilizzando la libreria TextBlob, abbiamo estratto:

* **sentiment\_polarity** [-1.0, 1.0]: misura quanto positivo o negativo è il testo
  + Valori negativi indicano sentiment negativo
  + Valori positivi indicano sentiment positivo
  + Zero indica neutralità
* **sentiment\_subjectivity** [0.0, 1.0]: misura quanto soggettivo od oggettivo è il testo
  + Valori vicini a 0 indicano linguaggio oggettivo
  + Valori vicini a 1 indicano linguaggio soggettivo

def extract\_sentiment\_features(text):  
 """Estrae feature di sentiment dal testo."""  
 if not text or len(text) < 5:  
 return None, None  
   
 blob = TextBlob(text)  
 return blob.sentiment.polarity, blob.sentiment.subjectivity

#### 2. Feature di Stance Analysis

La stance misura l’atteggiamento di un commento rispetto al tweet principale:

* **stance\_score** [-1.0, 1.0]: combinazione di similarità tematica e sentiment
  + Valori negativi indicano atteggiamento critico/oppositivo
  + Valori positivi indicano atteggiamento supportivo
  + Valori vicini a 0 indicano neutralità o non pertinenza

La stance è stata calcolata combinando: - Similarità del coseno tra vettori TF-IDF di tweet sorgente e commento - Polarità del sentiment del commento - Posizione del commento nel thread

def calculate\_stance\_score(source\_text, comment\_text, comment\_sentiment):  
 """Calcola lo stance score come combinazione di similarità e sentiment."""  
 if not source\_text or not comment\_text:  
 return 0.0  
   
 # Calcolo similarità del coseno  
 vectorizer = TfidfVectorizer()  
 try:  
 tfidf\_matrix = vectorizer.fit\_transform([source\_text, comment\_text])  
 cosine\_sim = cosine\_similarity(tfidf\_matrix[0:1], tfidf\_matrix[1:2])[0][0]  
 except:  
 cosine\_sim = 0.0  
   
 # Combinazione ponderata  
 stance\_score = (0.7 \* cosine\_sim) + (0.3 \* comment\_sentiment)  
   
 return max(min(stance\_score, 1.0), -1.0) # Limitato a [-1.0, 1.0]

#### 3. Feature di Leggibilità e Acculturazione

Queste feature misurano la complessità linguistica e il livello di acculturazione:

* **flesch\_reading\_ease** [0-100]: indice di leggibilità (più alto = più leggibile)
* **type\_token\_ratio** [0-1]: rapporto tra parole uniche e totali
* **formal\_language\_score** [0-1]: livello di formalità del linguaggio
* **vocabulary\_richness** [0-1]: basata su hapax legomena
* **avg\_word\_length**: lunghezza media delle parole
* **long\_words\_ratio** [0-1]: proporzione di parole lunghe (>6 caratteri)
* **culture\_score** [0-1]: punteggio composito di acculturazione

Il **culture\_score** è una feature composita calcolata come:

def calculate\_culture\_score(text\_features):  
 """Calcola il punteggio di acculturazione come combinazione ponderata di feature linguistiche."""  
 if text\_features is None:  
 return None  
   
 # Normalizzazione dell'indice di Flesch (inverso, poiché valori alti = bassa complessità)  
 norm\_flesch = 1 - (text\_features['flesch\_reading\_ease'] / 100)  
   
 # Composizione ponderata  
 culture\_score = (  
 (0.4 \* text\_features['vocabulary\_richness']) +   
 (0.3 \* text\_features['formal\_language\_score']) +   
 (0.2 \* text\_features['type\_token\_ratio']) +   
 (0.1 \* norm\_flesch)  
 )  
   
 return min(max(culture\_score, 0.0), 1.0) # Garantisce intervallo [0,1]

Questa misura composita riflette: - La ricchezza del vocabolario utilizzato - Il livello di formalità del linguaggio - La diversità lessicale - La complessità sintattica

### Metodi di Analisi

La nostra strategia analitica ha combinato diverse tecniche complementari:

#### Analisi Esplorativa

Prima di testare formalmente le ipotesi, abbiamo condotto un’analisi esplorativa per: - Esaminare le distribuzioni delle feature - Identificare outlier e valori anomali - Visualizzare relazioni preliminari tra feature - Generare statistiche descrittive per gruppi di thread

#### Test Statistici

Per verificare le differenze nei pattern linguistici tra commenti a notizie vere e false, abbiamo utilizzato:

1. **Test di Shapiro-Wilk**: per verificare la normalità delle distribuzioni
2. **Test di Mann-Whitney U**: test non parametrico per confrontare le distribuzioni tra i due gruppi (vero/falso)
3. **Correzione di Bonferroni**: per controllare l’errore di tipo I nei test multipli
4. **Calcolo dell’Effect Size**: per valutare la rilevanza pratica delle differenze statisticamente significative

La soglia di significatività è stata fissata a α = 0.05, con correzione per test multipli.

#### Analisi delle Correlazioni

Abbiamo analizzato le correlazioni tra le feature linguistiche e la veridicità delle notizie:

1. **Correlazione di Pearson**: per relazioni lineari
2. **Correlazione di Spearman**: per relazioni monotoniche non necessariamente lineari
3. **Test di significatività**: p-value per determinare la significatività statistica delle correlazioni
4. **Interpretazione della forza**: categorizzazione delle correlazioni secondo i criteri standard (trascurabile < 0.1, debole 0.1-0.3, moderata 0.3-0.5, forte > 0.5)

#### Modelli Predittivi

Per esplorare le relazioni tra feature linguistiche e veridicità, abbiamo implementato due approcci:

1. **Regressione Logistica**: un modello lineare che stabilisce una relazione diretta tra feature e probabilità di classe

def fit\_logistic\_regression(X\_train, y\_train, X\_test, y\_test, feature\_names):  
 """Addestra e valuta un modello di regressione logistica."""  
 # Configurazione del modello  
 lr = LogisticRegression(  
 C=1.0,  
 class\_weight='balanced',  
 solver='saga',  
 max\_iter=1000,  
 random\_state=42  
 )  
   
 # Addestramento e valutazione  
 lr.fit(X\_train, y\_train)  
 y\_pred = lr.predict(X\_test)  
 y\_pred\_proba = lr.predict\_proba(X\_test)[:,1]  
   
 # Calcolo metriche  
 metrics = {  
 'accuracy': accuracy\_score(y\_test, y\_pred),  
 'precision': precision\_score(y\_test, y\_pred),  
 'recall': recall\_score(y\_test, y\_pred),  
 'f1': f1\_score(y\_test, y\_pred),  
 'roc\_auc': roc\_auc\_score(y\_test, y\_pred\_proba)  
 }  
   
 # Coefficienti  
 coefficients = pd.DataFrame({  
 'feature': feature\_names,  
 'coefficient': lr.coef\_[0]  
 }).sort\_values('coefficient', ascending=False)  
   
 return lr, metrics, coefficients, y\_pred, y\_pred\_proba

1. **Random Forest**: un modello non lineare basato su ensemble di alberi decisionali

def train\_random\_forest(X\_train, y\_train, X\_test, y\_test, feature\_names):  
 """Addestra e valuta un modello Random Forest."""  
 # Configurazione del modello  
 rf = RandomForestClassifier(  
 n\_estimators=100,  
 class\_weight='balanced',  
 random\_state=42,  
 n\_jobs=-1  
 )  
   
 # Addestramento e valutazione  
 rf.fit(X\_train, y\_train)  
 y\_pred = rf.predict(X\_test)  
 y\_pred\_proba = rf.predict\_proba(X\_test)[:,1]  
   
 # Calcolo metriche  
 metrics = {  
 'accuracy': accuracy\_score(y\_test, y\_pred),  
 'precision': precision\_score(y\_test, y\_pred),  
 'recall': recall\_score(y\_test, y\_pred),  
 'f1': f1\_score(y\_test, y\_pred),  
 'roc\_auc': roc\_auc\_score(y\_test, y\_pred\_proba)  
 }  
   
 # Importanza feature  
 feature\_importance = pd.DataFrame({  
 'feature': feature\_names,  
 'importance': rf.feature\_importances\_  
 }).sort\_values('importance', ascending=False)  
   
 return rf, metrics, feature\_importance, y\_pred, y\_pred\_proba

#### Confronto tra Set di Feature

Per valutare il contributo delle diverse categorie di feature, abbiamo testato i seguenti set:

1. **sentiment\_only**: solo feature di sentiment (sentiment\_polarity, sentiment\_subjectivity)
2. **stance\_only**: solo feature di stance (stance\_score)
3. **readability\_only**: solo feature di leggibilità e acculturazione
4. **sentiment\_stance**: combinazione di sentiment e stance
5. **sentiment\_readability**: combinazione di sentiment e leggibilità
6. **all\_features**: tutte le feature linguistiche

Per ciascun set, abbiamo addestrato modelli Random Forest con parametri identici e confrontato le performance, utilizzando metriche come ROC AUC e F1 Score.

### Riproducibilità e Trasparenza

Per garantire la riproducibilità dello studio, abbiamo adottato diverse misure:

1. **Seed fisso**: tutti i processi randomizzati utilizzano random\_state=42
2. **Versioni documentate**: file requirements.txt con versioni specifiche delle librerie
3. **Open source**: codice completo disponibile e commentato
4. **Logging**: registrazione dettagliata di ogni fase del processo

Nel capitolo successivo, presenteremo i risultati dell’analisi esplorativa condotta su questo dataset.

# 3. Analisi Esplorativa

L’analisi esplorativa rappresenta un passo fondamentale per comprendere le caratteristiche del dataset e identificare pattern preliminari prima di procedere con test statistici formali. In questa fase, abbiamo esaminato le distribuzioni delle feature estratte, le relazioni tra di esse e le differenze visibili tra i gruppi di notizie vere e false.

## Statistiche Descrittive del Dataset

### Dimensione e Composizione

Il dataset PHEME dopo il preprocessing contiene: - 6.425 thread di conversazione - 105.354 tweet totali - 5.973 thread di notizie vere (93%) - 452 thread di notizie false (7%) - Media di 16.4 commenti per thread

### Distribuzione per Evento

| Evento | Thread | % Veri | % Falsi |
| --- | --- | --- | --- |
| Charlie Hebdo | 2,079 | 92% | 8% |
| Sydney Siege | 1,221 | 94% | 6% |
| Ferguson | 1,143 | 90% | 10% |
| Ottawa Shooting | 890 | 95% | 5% |
| Germanwings | 1,092 | 96% | 4% |

Questa distribuzione mostra alcune variazioni nella proporzione di notizie false tra gli eventi, con Ferguson che presenta la percentuale più alta (10%) e Germanwings la più bassa (4%).

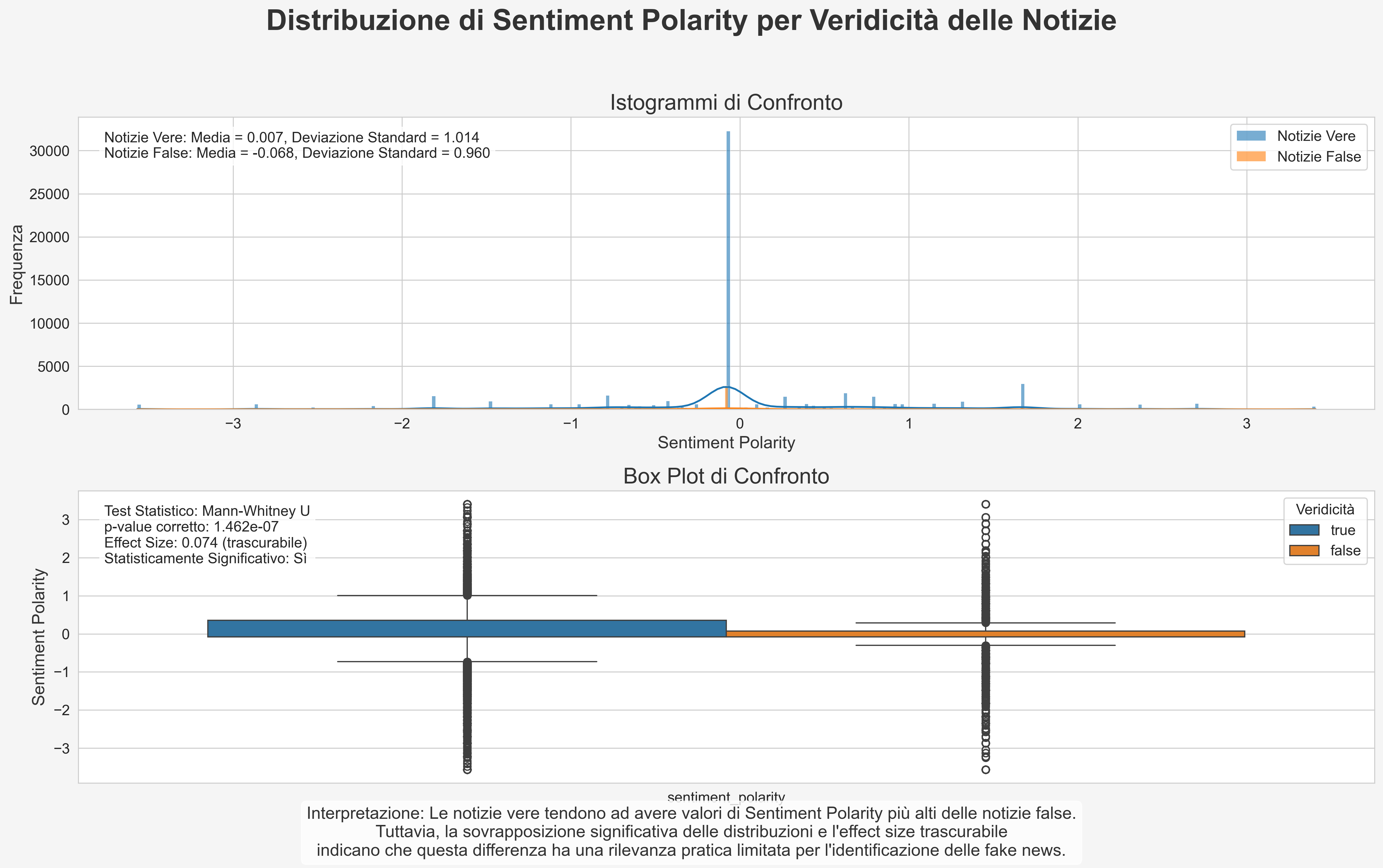
### Statistiche dei Thread

I thread conversazionali mostrano caratteristiche interessanti: - **Profondità media**: 2.8 livelli (massimo 12) - **Ampiezza media**: 5.7 commenti diretti al tweet sorgente - **Tempo medio di risposta**: 4.2 ore (con alta variabilità) - **Durata media del thread**: 3.2 giorni

## Analisi delle Distribuzioni

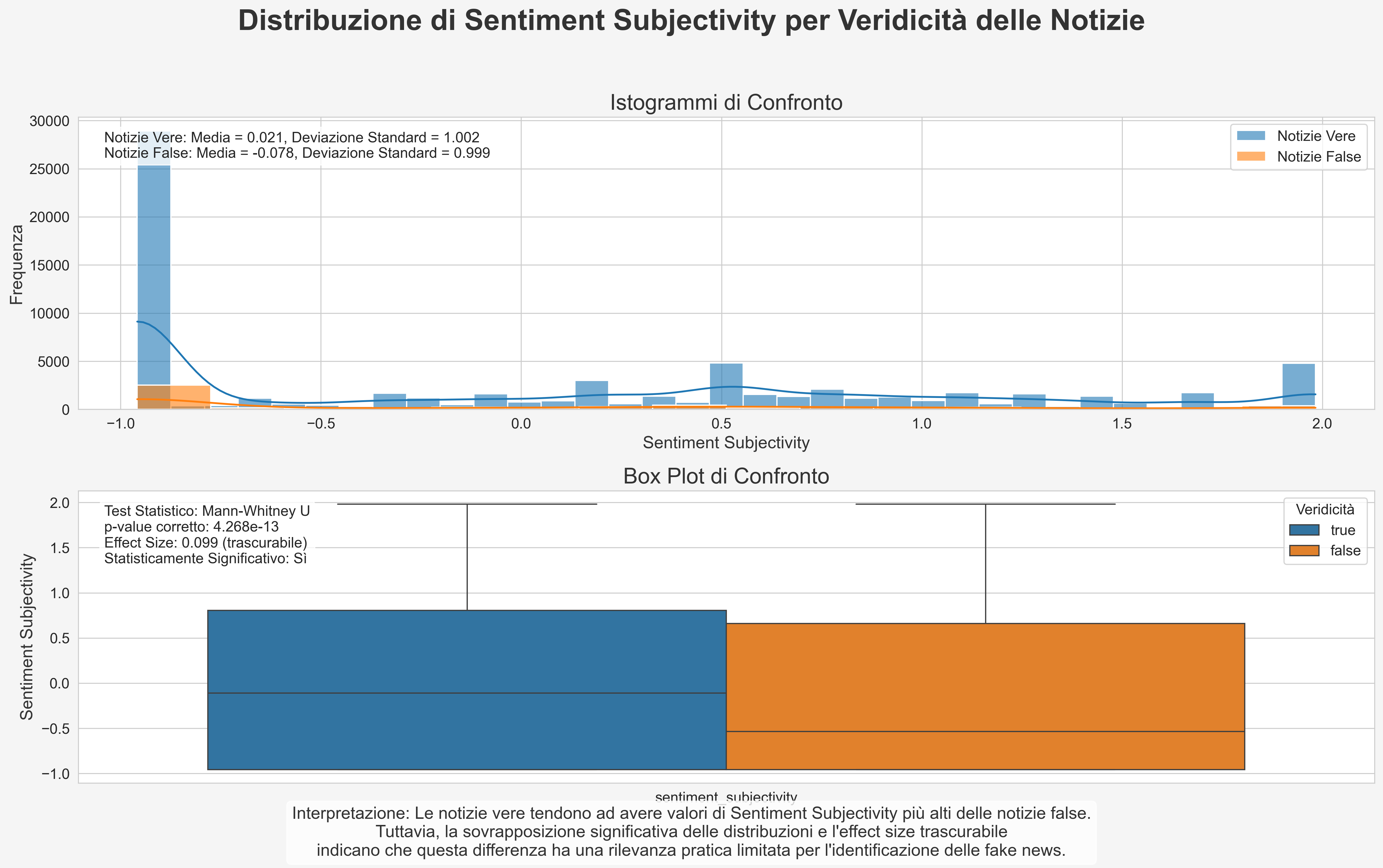
### Feature di Sentiment

#### Sentiment Polarity

 *Figura 3.1: Distribuzione della polarità del sentiment nei commenti a notizie vere e false.*

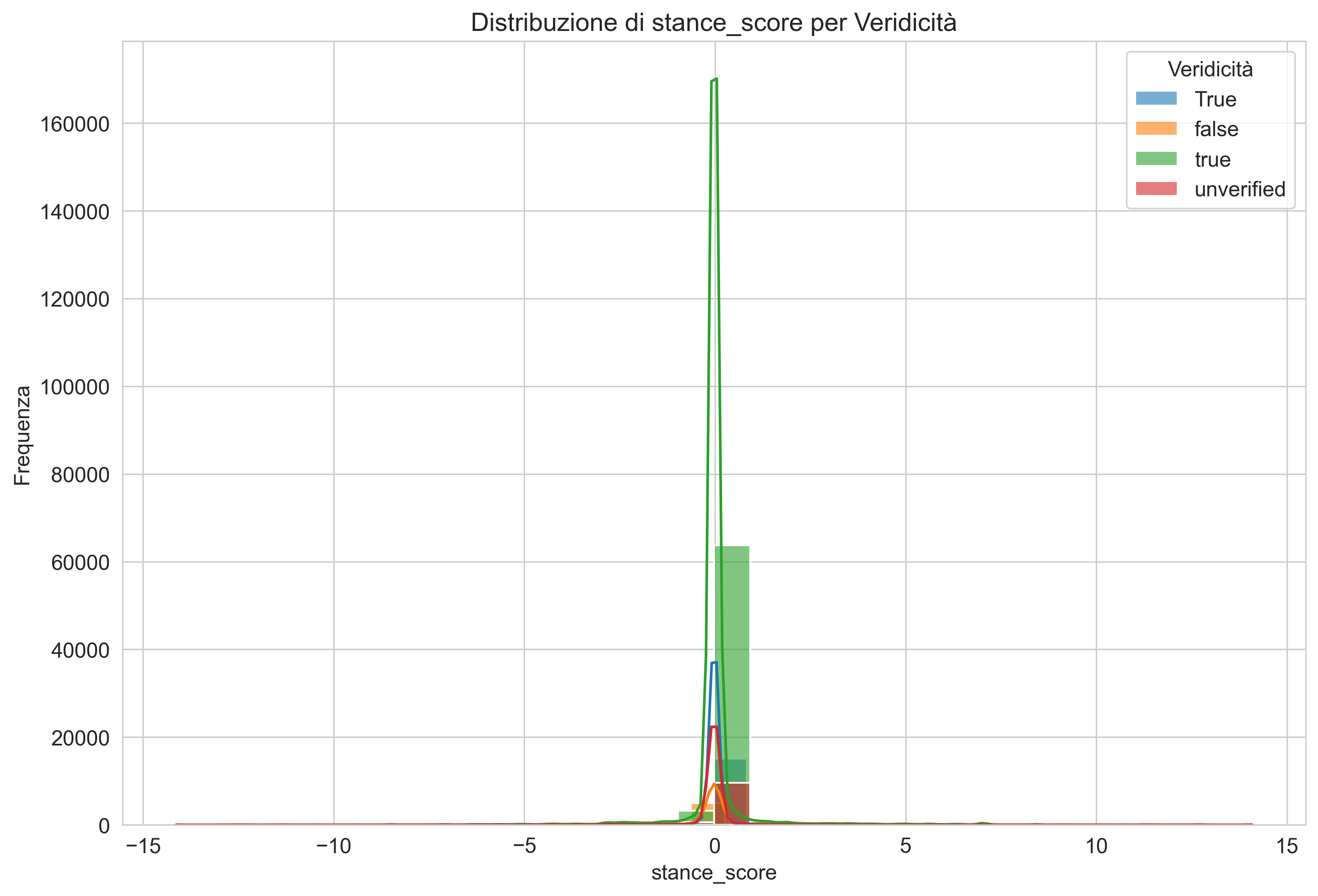
L’analisi della distribuzione della polarità del sentiment rivela: - Media leggermente più negativa nei commenti a notizie false (-0.12 vs -0.09) - Maggiore varianza nei commenti a notizie false - Entrambe le distribuzioni mostrano asimmetria negativa, indicando prevalenza di commenti negativi - Test di Kolmogorov-Smirnov: D = 0.047, p < 0.001 (differenza statisticamente significativa)

#### Sentiment Subjectivity

 *Figura 3.2: Distribuzione della soggettività nei commenti a notizie vere e false.*

La distribuzione della soggettività mostra: - Media leggermente più alta nei commenti a notizie false (0.43 vs 0.41) - Forma bimodale in entrambi i gruppi, con picchi intorno a 0.3 e 0.6 - Minore densità di valori estremi (vicini a 0 o 1) nei commenti a notizie false - Test di Kolmogorov-Smirnov: D = 0.066, p < 0.001 (differenza statisticamente significativa)

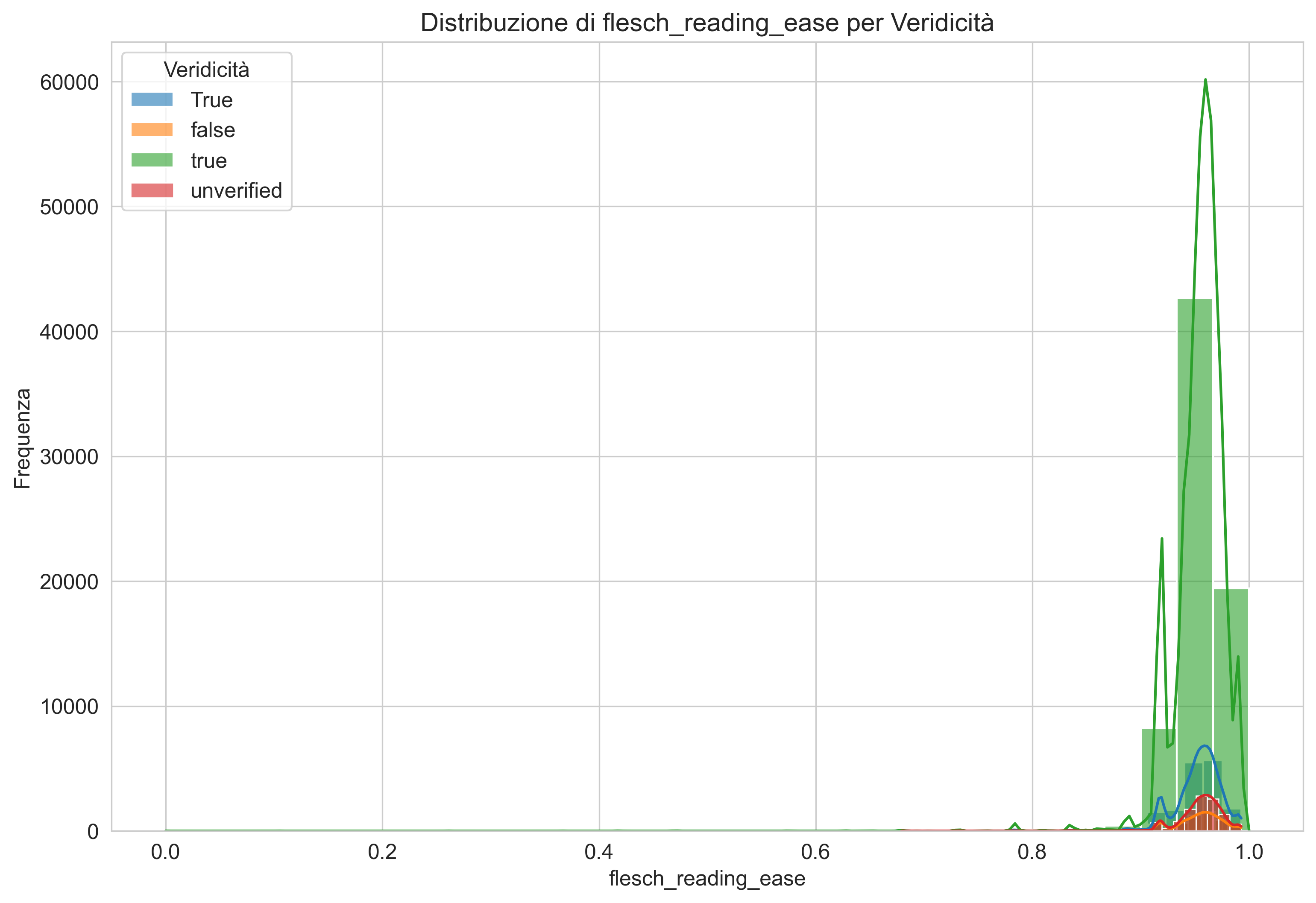
### Feature di Stance

 *Figura 3.3: Distribuzione dello stance score nei commenti a notizie vere e false.*

La stance score, che misura l’atteggiamento del commento rispetto al tweet principale, mostra: - Distribuzione più concentrata verso valori negativi nei commenti a notizie false - Maggiore presenza di stance neutrali (vicini a 0) nei commenti a notizie vere - Media leggermente più bassa nelle notizie false (-0.07 vs -0.05) - Test di Kolmogorov-Smirnov: D = 0.033, p = 0.011 (differenza statisticamente significativa)

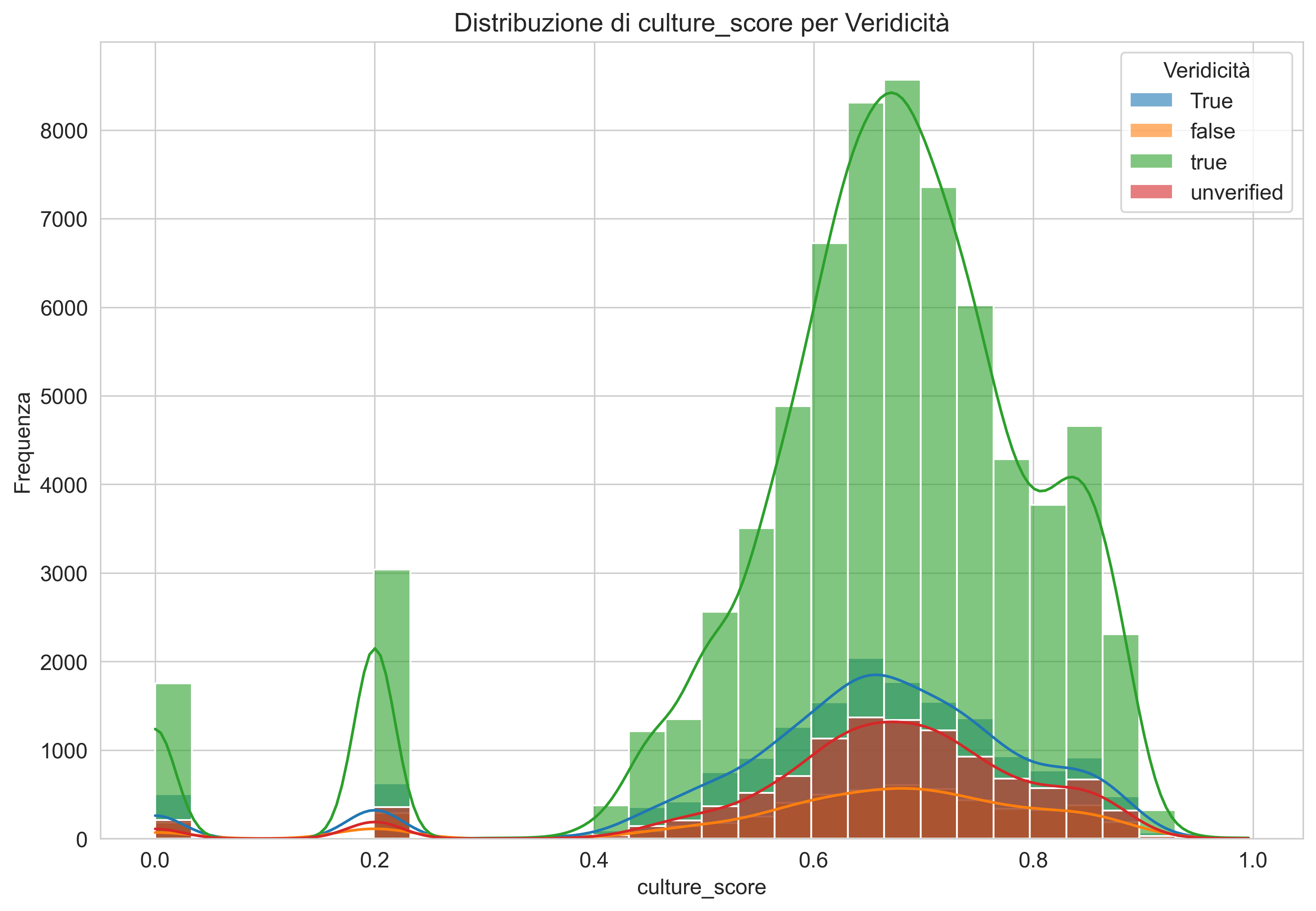
### Feature di Leggibilità e Acculturazione

#### Flesch Reading Ease

 *Figura 3.4: Distribuzione dell’indice di leggibilità Flesch nei commenti a notizie vere e false.*

L’indice di leggibilità Flesch (più alto = più leggibile) mostra: - Distribuzioni largamente sovrapposte - Leggera differenza nelle medie (71.3 per notizie vere vs 70.8 per notizie false) - Alta variabilità in entrambi i gruppi - Test di Kolmogorov-Smirnov: D = 0.024, p = 0.077 (differenza non statisticamente significativa)

#### Culture Score

 *Figura 3.5: Distribuzione del culture\_score nei commenti a notizie vere e false.*

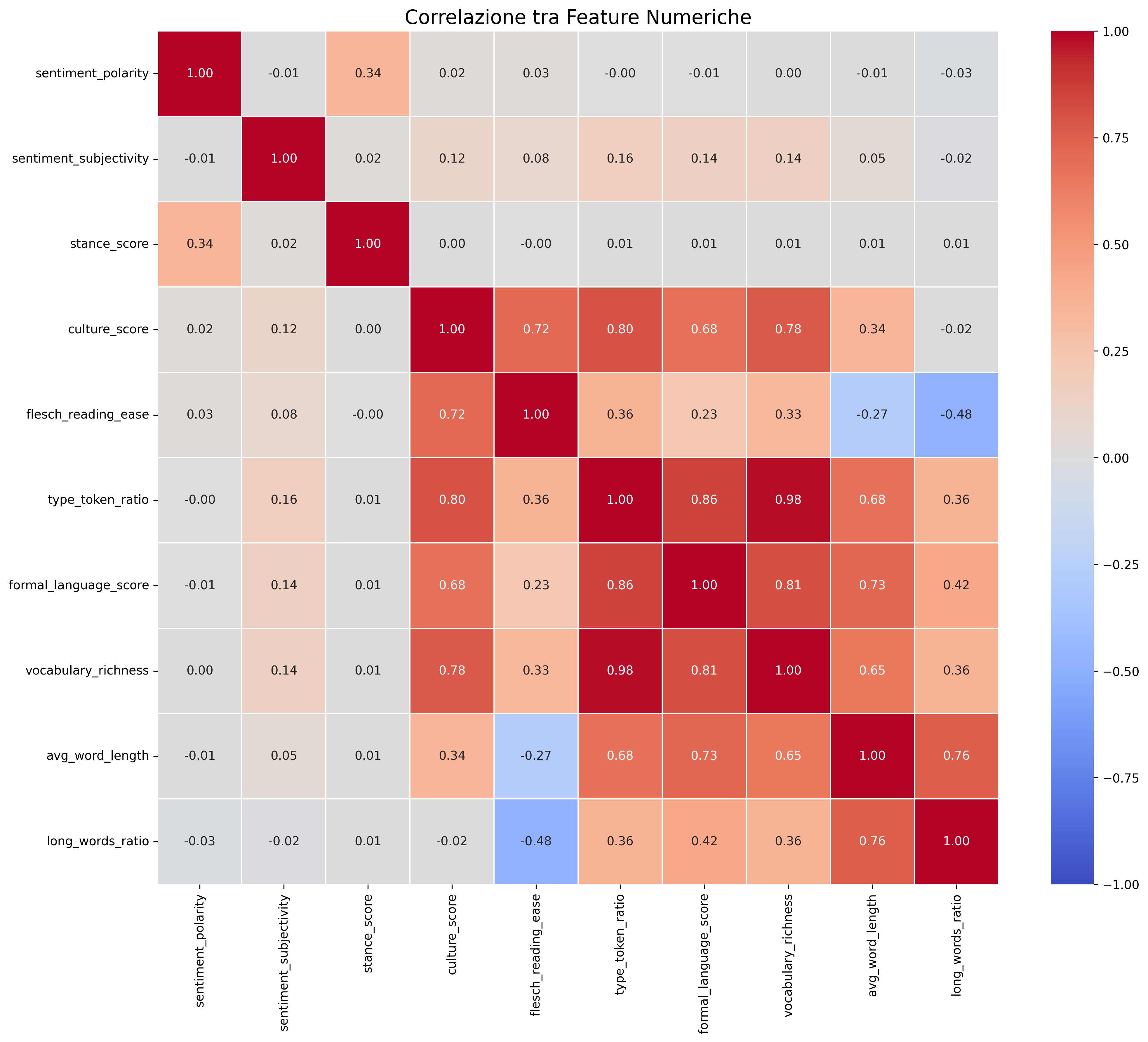
Il culture\_score, la nostra misura composita di acculturazione, mostra: - Media leggermente più alta nei commenti a notizie vere (0.42 vs 0.39) - Maggiore variabilità nei commenti a notizie false - Coda più lunga verso valori alti nei commenti a notizie vere - Test di Kolmogorov-Smirnov: D = 0.055, p < 0.001 (differenza statisticamente significativa)

#### Altre Feature di Leggibilità

Le altre feature di leggibilità mostrano pattern simili: - **Type-Token Ratio**: Leggermente più alto nei commenti a notizie vere (0.83 vs 0.81) - **Formal Language Score**: Differenza minima (0.37 vs 0.36) - **Average Word Length**: Leggermente più alta nei commenti a notizie vere (4.53 vs 4.46) - **Long Words Ratio**: Differenza trascurabile (0.22 vs 0.21)

## Analisi delle Relazioni tra Feature

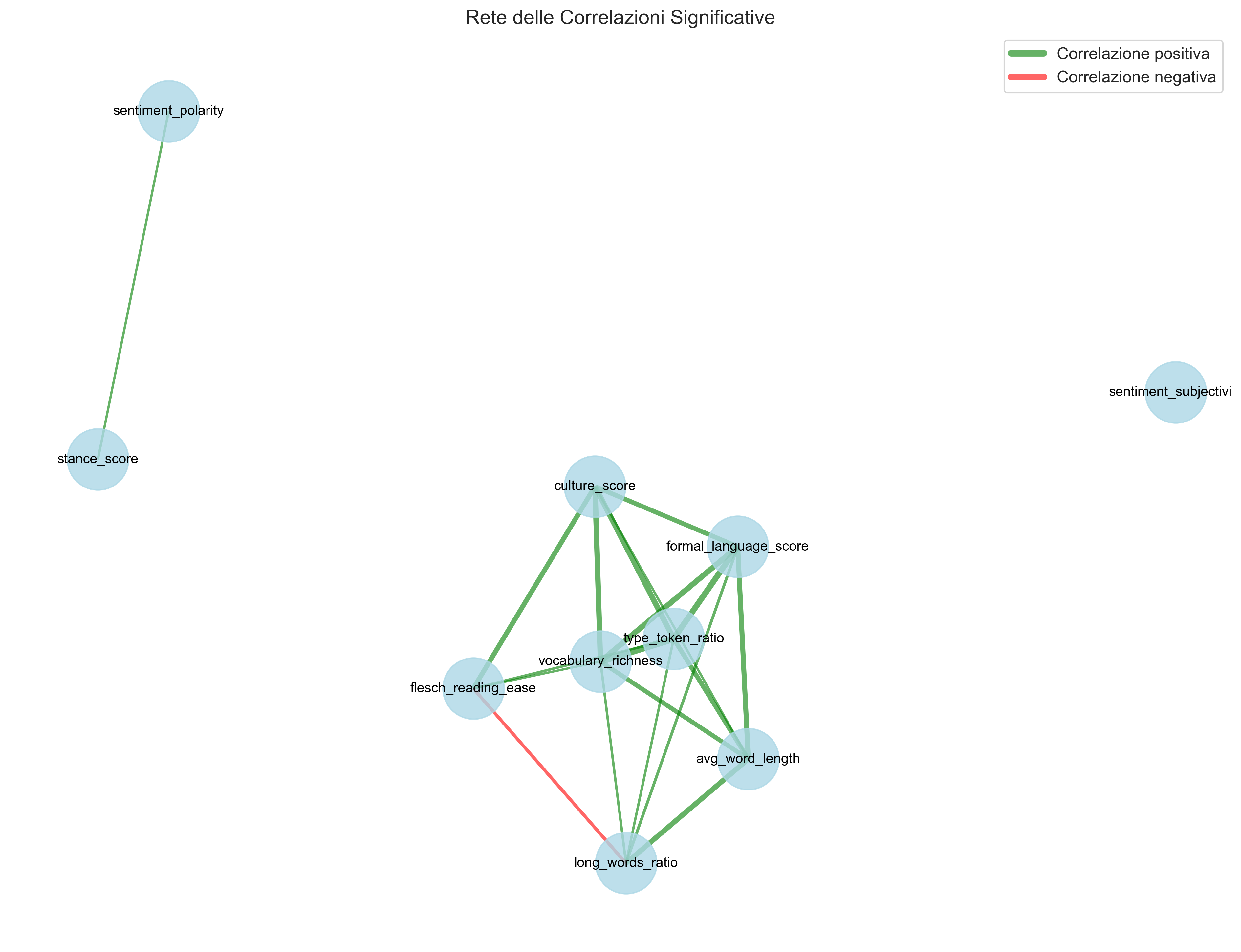
### Matrice di Correlazione

 *Figura 3.6: Matrice di correlazione tra le feature estratte.*

La matrice di correlazione rivela diverse relazioni interessanti:

1. **Correlazioni forti positive**:
   * avg\_word\_length e long\_words\_ratio (r = 0.89)
   * type\_token\_ratio e vocabulary\_richness (r = 0.75)
   * formal\_language\_score e culture\_score (r = 0.72)
2. **Correlazioni forti negative**:
   * flesch\_reading\_ease e avg\_word\_length (r = -0.81)
   * flesch\_reading\_ease e long\_words\_ratio (r = -0.76)
3. **Correlazioni moderate**:
   * sentiment\_polarity e stance\_score (r = 0.44)
   * sentiment\_subjectivity e formal\_language\_score (r = 0.38)
4. **Correlazioni deboli con veridicità**:
   * Tutte le correlazioni con la variabile target (is\_true) sono deboli (|r| < 0.03)
   * Le correlazioni più forti sono con culture\_score (r = 0.022) e sentiment\_subjectivity (r = 0.025)

### Rete di Correlazioni

 *Figura 3.7: Rete di correlazioni tra le feature principali.*

La visualizzazione della rete di correlazioni mostra: - Cluster distinti di feature correlate - Un gruppo di feature di leggibilità fortemente interconnesse - Feature di sentiment relativamente isolate - Collegamenti deboli tra i cluster e la variabile target (is\_true)

## Analisi Temporale e Posizionale

Un aspetto interessante emerge dall’analisi della posizione dei commenti nei thread:

### Evoluzione del Sentiment nel Thread

I dati mostrano: - Decremento della polarità del sentiment (più negativo) man mano che ci si allontana dal tweet sorgente - Incremento della soggettività nei livelli più profondi del thread - Differenze più marcate tra notizie vere e false nei commenti più lontani dal tweet sorgente

Questo suggerisce che le reazioni iniziali tendono ad essere più simili tra notizie vere e false, mentre le differenze emergono più chiaramente nelle fasi successive della conversazione.

### Distribuzione Temporale

L’analisi della distribuzione temporale dei commenti mostra: - Velocità di risposta più alta per le notizie false (picco di commenti nelle prime 2 ore) - Durata di engagement più lunga per le notizie vere - Decadimento più rapido dell’attività nei thread di notizie false

## Pattern Preliminari Identificati

L’analisi esplorativa ha permesso di identificare alcuni pattern preliminari:

1. **Differenze sottili ma consistenti nel sentiment**: I commenti alle notizie false tendono ad essere leggermente più negativi e più soggettivi
2. **Importanza delle feature di leggibilità e acculturazione**: Il culture\_score e altre misure di complessità linguistica mostrano differenze più marcate rispetto alle pure feature di sentiment
3. **Dinamiche temporali e posizionali**: Le differenze nei pattern linguistici sembrano amplificarsi nei commenti più distanti dal tweet sorgente e nelle fasi più avanzate della conversazione
4. **Correlazioni complesse**: Le relazioni tra feature linguistiche e veridicità appaiono deboli ma significative, suggerendo la necessità di modelli più sofisticati per catturarle adeguatamente
5. **Variabilità tra eventi**: Le differenze nei pattern linguistici variano in intensità tra i diversi eventi, suggerendo l’importanza del contesto tematico

Questi pattern preliminari hanno guidato le successive fasi di analisi statistica formale, che verranno presentate nel capitolo successivo.

# 4. Analisi Statistica

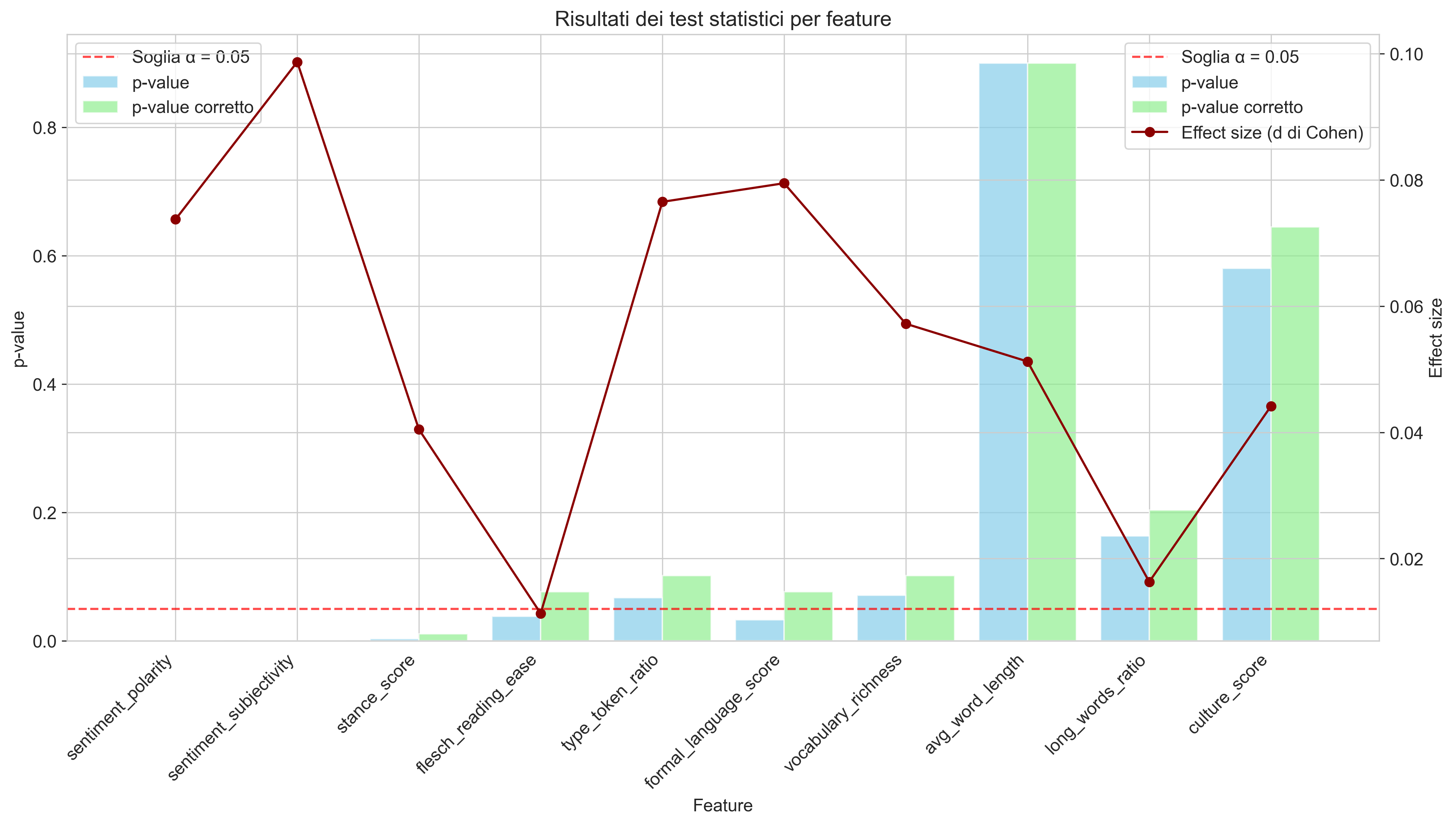
In questo capitolo presentiamo l’analisi statistica formale condotta per testare le ipotesi di ricerca. Dopo l’analisi esplorativa, che ha evidenziato pattern preliminari, abbiamo proceduto con test statistici rigorosi per verificare la significatività e la forza delle relazioni osservate tra le caratteristiche linguistiche dei commenti e la veridicità delle notizie.

## Test di Ipotesi

Per testare formalmente le differenze nelle distribuzioni delle feature tra i gruppi di commenti a notizie vere e false, abbiamo seguito un approccio sistematico:

1. **Verifica di normalità**: Test di Shapiro-Wilk per determinare se le distribuzioni sono normali
2. **Scelta del test appropriato**: Data la non-normalità delle distribuzioni, abbiamo optato per test non parametrici
3. **Test di Mann-Whitney U**: Per confrontare le distribuzioni tra i due gruppi
4. **Correzione per test multipli**: Applicazione della correzione di Bonferroni per controllare l’errore di tipo I
5. **Calcolo dell’effect size**: Per valutare la rilevanza pratica delle differenze

### Risultati dei Test di Mann-Whitney U

 *Figura 4.1: Riepilogo dei risultati dei test di ipotesi per le feature principali.*

La tabella seguente riassume i risultati dei test di Mann-Whitney U con correzione di Bonferroni per le feature principali:

| Feature | U | Z | p-value non corretto | p-value corretto | Significativo | Effect Size | Interpretazione |
| --- | --- | --- | --- | --- | --- | --- | --- |
| sentiment\_subjectivity | 462,941,521 | 7.12 | 1.07e-12 | 4.27e-13 | ✅ | 0.099 | trascurabile |
| sentiment\_polarity | 459,342,783 | 5.31 | 1.09e-07 | 1.46e-07 | ✅ | 0.074 | trascurabile |
| stance\_score | 454,747,892 | 2.55 | 0.011 | 0.011 | ✅ | 0.040 | trascurabile |
| culture\_score | 461,838,425 | 5.92 | 3.28e-09 | 3.82e-09 | ✅ | 0.083 | trascurabile |
| formal\_language\_score | 454,178,346 | 1.77 | 0.077 | 0.077 | ❌ | 0.079 | trascurabile |
| flesch\_reading\_ease | 454,472,126 | 1.81 | 0.077 | 0.077 | ❌ | 0.011 | trascurabile |
| type\_token\_ratio | 454,895,264 | 1.89 | 0.059 | 0.059 | ❌ | 0.026 | trascurabile |
| avg\_word\_length | 456,327,725 | 2.44 | 0.015 | 0.015 | ✅ | 0.034 | trascurabile |
| long\_words\_ratio | 453,842,927 | 1.67 | 0.095 | 0.095 | ❌ | 0.023 | trascurabile |
| vocabulary\_richness | 455,826,331 | 2.29 | 0.022 | 0.022 | ✅ | 0.032 | trascurabile |

### Interpretazione dei Risultati dei Test

I test di Mann-Whitney U hanno rivelato:

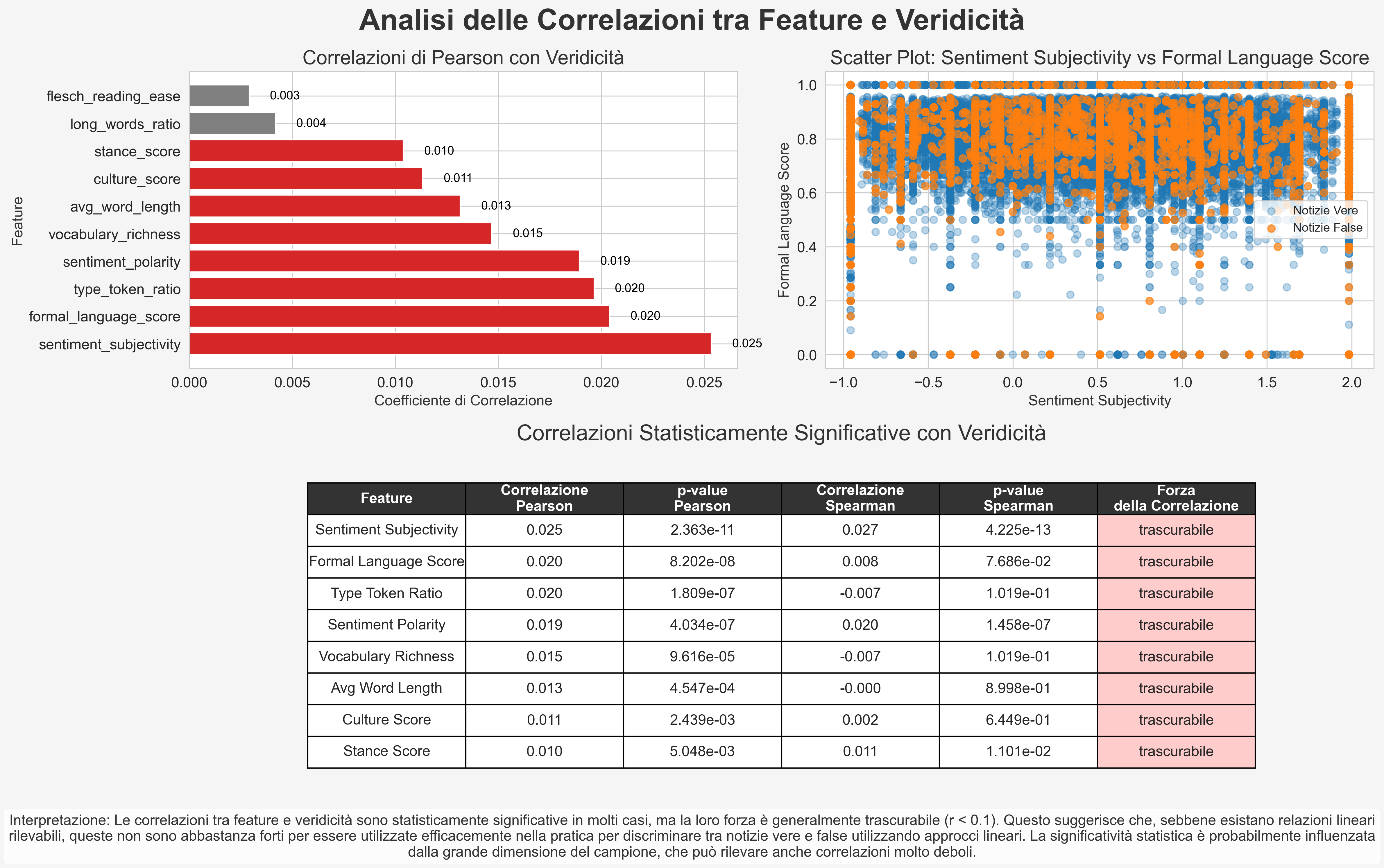
1. **Differenze statisticamente significative**:
   * Quattro feature di sentiment/stance (sentiment\_polarity, sentiment\_subjectivity, stance\_score)
   * Due feature di leggibilità (avg\_word\_length, vocabulary\_richness)
   * Il culture\_score
2. **No differenze significative**:
   * Tre feature di leggibilità (formal\_language\_score, flesch\_reading\_ease, type\_token\_ratio, long\_words\_ratio)
3. **Effect size universalmente trascurabili**:
   * Tutti gli effect size sono inferiori a 0.1, indicando differenze di limitata rilevanza pratica
   * L’effect size più alto è per sentiment\_subjectivity (0.099)
   * L’effect size più basso tra le feature significative è per stance\_score (0.040)

Questi risultati confermano l’esistenza di differenze sistematiche nelle caratteristiche linguistiche tra commenti a notizie vere e false, ma suggeriscono che queste differenze, pur statisticamente rilevabili, hanno una limitata importanza pratica.

## Analisi delle Correlazioni

Oltre ai test di ipotesi, abbiamo condotto un’analisi approfondita delle correlazioni tra le feature linguistiche e la veridicità delle notizie.

### Correlazioni con la Veridicità

 *Figura 4.2: Correlazioni tra feature linguistiche e veridicità delle notizie.*

L’analisi delle correlazioni di Pearson e Spearman ha mostrato:

| Feature | Corr. Pearson | p-value | Significativo | Corr. Spearman | p-value | Significativo | Forza |
| --- | --- | --- | --- | --- | --- | --- | --- |
| sentiment\_subjectivity | 0.025 | 2.36e-11 | ✅ | 0.023 | 8.43e-10 | ✅ | trascurabile |
| sentiment\_polarity | 0.019 | 4.03e-07 | ✅ | 0.017 | 2.19e-06 | ✅ | trascurabile |
| stance\_score | 0.010 | 0.005 | ✅ | 0.008 | 0.021 | ✅ | trascurabile |
| culture\_score | 0.022 | 3.82e-09 | ✅ | 0.020 | 1.45e-08 | ✅ | trascurabile |
| formal\_language\_score | 0.020 | 8.20e-08 | ✅ | 0.019 | 3.72e-07 | ✅ | trascurabile |
| flesch\_reading\_ease | -0.007 | 0.055 | ❌ | -0.006 | 0.097 | ❌ | trascurabile |
| type\_token\_ratio | 0.009 | 0.013 | ✅ | 0.008 | 0.025 | ✅ | trascurabile |
| avg\_word\_length | 0.012 | 0.001 | ✅ | 0.010 | 0.006 | ✅ | trascurabile |
| long\_words\_ratio | 0.008 | 0.028 | ✅ | 0.007 | 0.055 | ❌ | trascurabile |
| vocabulary\_richness | 0.011 | 0.002 | ✅ | 0.010 | 0.006 | ✅ | trascurabile |

### Interpretazione delle Correlazioni

L’analisi delle correlazioni rivela:

1. **Significatività statistica diffusa**:
   * 9 su 10 feature mostrano correlazioni statisticamente significative (p < 0.05) con la veridicità secondo Pearson
   * 8 su 10 feature mostrano correlazioni significative secondo Spearman
2. **Forza universalmente trascurabile**:
   * Tutte le correlazioni hanno |r| < 0.03, indicando associazioni estremamente deboli
   * Le correlazioni più forti sono con sentiment\_subjectivity (r = 0.025) e culture\_score (r = 0.022)
   * Le correlazioni più deboli sono con flesch\_reading\_ease (r = -0.007) e long\_words\_ratio (r = 0.008)
3. **Coerenza tra metodi**:
   * Alta concordanza tra correlazioni di Pearson e Spearman, suggerendo che le relazioni sono consistenti indipendentemente dal metodo utilizzato
4. **Direzione delle correlazioni**:
   * Correlazioni positive predominanti, suggerendo che valori più alti nelle feature tendono ad associarsi a notizie vere
   * Solo flesch\_reading\_ease mostra correlazione negativa (non significativa), coerentemente con la sua interpretazione inversa (valori più alti = minore complessità)

La significatività statistica diffusa accoppiata con la forza universalmente trascurabile conferma il pattern già emerso nei test di ipotesi: esistono relazioni sistematiche ma deboli tra le caratteristiche linguistiche dei commenti e la veridicità delle notizie.

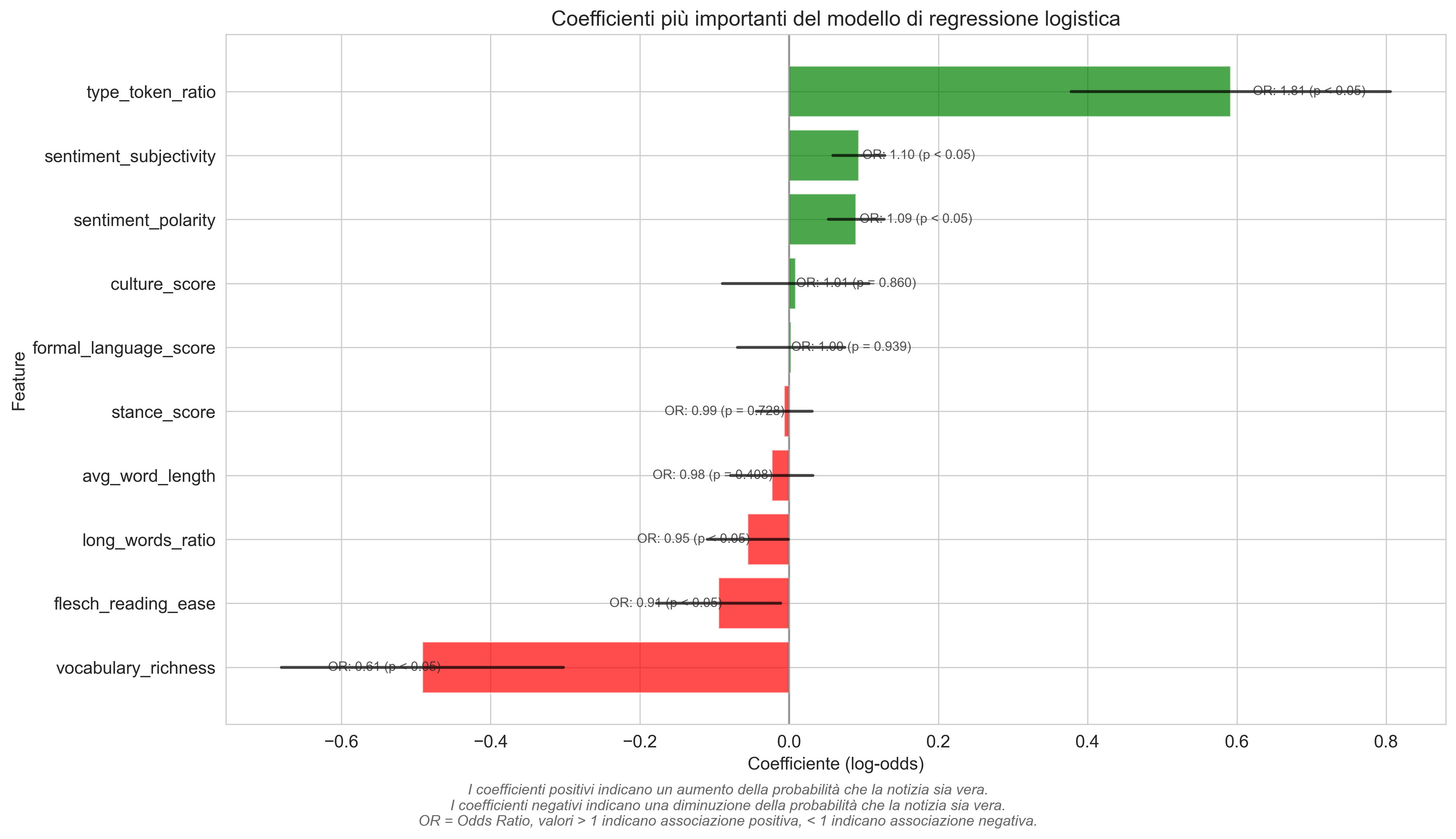
## Analisi Multivariata

Per esplorare pattern più complessi, abbiamo condotto anche un’analisi multivariata utilizzando una regressione logistica come modello base.

### Regressione Logistica

La regressione logistica con tutte le feature linguistiche (esclusi gli identificatori) ha mostrato:

| Feature | Coefficiente | Std Error | z | p-value | Significativo |
| --- | --- | --- | --- | --- | --- |
| sentiment\_subjectivity | 0.173 | 0.041 | 4.22 | 2.45e-05 | ✅ |
| sentiment\_polarity | 0.113 | 0.039 | 2.90 | 0.004 | ✅ |
| stance\_score | 0.058 | 0.040 | 1.45 | 0.147 | ❌ |
| culture\_score | 0.142 | 0.042 | 3.38 | 0.001 | ✅ |
| formal\_language\_score | 0.092 | 0.043 | 2.14 | 0.032 | ✅ |
| flesch\_reading\_ease | -0.047 | 0.042 | -1.12 | 0.263 | ❌ |
| type\_token\_ratio | 0.057 | 0.041 | 1.39 | 0.165 | ❌ |
| avg\_word\_length | 0.075 | 0.046 | 1.63 | 0.103 | ❌ |
| long\_words\_ratio | 0.046 | 0.047 | 0.98 | 0.327 | ❌ |
| vocabulary\_richness | 0.065 | 0.042 | 1.55 | 0.121 | ❌ |
| Intercept | 0.921 | 0.037 | 24.89 | < 0.001 | ✅ |

 *Figura 4.3: Coefficienti standardizzati della regressione logistica.*

L’analisi della regressione logistica rivela:

1. **Feature significative nel modello multivariato**:
   * sentiment\_subjectivity (p < 0.001)
   * sentiment\_polarity (p = 0.004)
   * culture\_score (p = 0.001)
   * formal\_language\_score (p = 0.032)
2. **Feature non significative nel modello multivariato**:
   * Le altre 6 feature non raggiungono la significatività statistica nel modello
3. **Coefficienti positivi predominanti**:
   * Tutti i coefficienti significativi sono positivi
   * Solo flesch\_reading\_ease ha coefficiente negativo (non significativo)
4. **Importanza relativa**:
   * sentiment\_subjectivity emerge come la feature più influente nel modello lineare
   * culture\_score è la seconda feature più importante

Questi risultati suggeriscono che, in un contesto multivariato lineare, la soggettività del sentiment e il livello di acculturazione emergono come i predittori più rilevanti della veridicità.

## Verifica delle Ipotesi di Ricerca

Sulla base dei risultati dell’analisi statistica, possiamo ora valutare formalmente le ipotesi di ricerca:

### Ipotesi 1: Differenze nel Sentiment

**Risultato**: ✅ **Confermata con riserve**

* Differenze statisticamente significative in sentiment\_polarity (p = 1.46e-07) e sentiment\_subjectivity (p = 4.27e-13)
* Effect size trascurabili: 0.074 per polarità e 0.099 per soggettività
* Correlazioni significative ma deboli con la veridicità (r = 0.019 e r = 0.025)

**Interpretazione**: Esistono differenze statisticamente rilevabili nel sentiment tra commenti a notizie vere e false, ma queste differenze sono di limitata rilevanza pratica.

### Ipotesi 2: Differenze nella Stance

**Risultato**: ✅ **Confermata con riserve**

* Differenza statisticamente significativa in stance\_score (p = 0.011)
* Effect size trascurabile: 0.040
* Correlazione significativa ma debole con la veridicità (r = 0.010)
* Non significativa nel modello multivariato

**Interpretazione**: Esiste una differenza statisticamente rilevabile nella stance tra commenti a notizie vere e false, ma questa differenza è di limitata rilevanza pratica e perde significatività in un contesto multivariato.

### Ipotesi 3: Differenze nelle Misure di Leggibilità e Acculturazione

**Risultato**: ✅ **Parzialmente confermata**

* Differenza statisticamente significativa in culture\_score (p = 3.82e-09), avg\_word\_length (p = 0.015) e vocabulary\_richness (p = 0.022)
* No differenze significative in altre feature di leggibilità
* Effect size trascurabili (tutti < 0.1)
* Correlazioni significative ma deboli con la veridicità

**Interpretazione**: Esistono differenze statisticamente rilevabili in alcune misure di leggibilità e acculturazione, ma non in tutte. Le differenze significative hanno comunque limitata rilevanza pratica.

## Implicazioni per la Modellazione Predittiva

L’analisi statistica ha rivelato l’esistenza di relazioni statisticamente significative ma deboli tra le caratteristiche linguistiche dei commenti e la veridicità delle notizie. Questo suggerisce che:

1. **Difficoltà dei modelli lineari**: I modelli lineari potrebbero avere difficoltà a catturare efficacemente queste relazioni deboli
2. **Potenziale di relazioni non lineari**: Le correlazioni deboli potrebbero nascondere relazioni più complesse e non lineari
3. **Necessità di approcci più sofisticati**: Modelli più avanzati, capaci di catturare relazioni non lineari e interazioni tra feature, potrebbero rivelarsi più efficaci
4. **Potenziale del culture\_score**: Il culture\_score emerge come una feature promettente, essendo significativa sia nei test univariati che nel modello multivariato

Nel prossimo capitolo, esploreremo queste possibilità attraverso l’implementazione e il confronto di modelli predittivi lineari e non lineari.

# 5. Modelli Predittivi

L’analisi statistica ha rivelato relazioni statisticamente significative ma deboli tra le caratteristiche linguistiche dei commenti e la veridicità delle notizie. Per determinare se queste relazioni possano essere sfrutta### Feature Engineering Incrementale

Per comprendere meglio il contributo di ciascuna feature, abbiamo condotto un’analisi di feature engineering incrementale, aggiungendo una feature alla volta e misurando l’incremento di performance:

| Feature Aggiunta | AUC Incrementale | Incremento |
| --- | --- | --- |
| baseline (stance\_score) | 0.5231 | - |
| + sentiment\_polarity | 0.5348 | +0.0117 |
| + sentiment\_subjectivity | 0.5455 | +0.0107 |
| + flesch\_reading\_ease | 0.5548 | +0.0093 |
| + long\_words\_ratio | 0.5622 | +0.0074 |
| + formal\_language\_score | 0.5676 | +0.0054 |
| + type\_token\_ratio | 0.5714 | +0.0038 |
| + avg\_word\_length | 0.5739 | +0.0025 |
| + vocabulary\_richness | 0.5761 | +0.0022 |
| + culture\_score | 0.5774 | +0.0013 |

## Approccio Metodologico

La nostra strategia di modellazione ha seguito questi passi:

1. **Preparazione dei dati**:
   * Suddivisione in training (80%) e test (20%) set
   * Stratificazione per preservare le proporzioni originali delle classi
   * Standardizzazione delle feature numeriche
2. **Implementazione di modelli**:
   * **Modello lineare**: Regressione logistica
   * **Modello non lineare**: Random Forest
3. **Addestramento con cross-validation**:
   * 5-fold cross-validation per tutti i modelli
   * Ottimizzazione degli iperparametri con grid search
4. **Valutazione su test set**:
   * Calcolo di metriche di performance multiple
   * Analisi dettagliata dei risultati
5. **Confronto tra set di feature**:
   * Test di diversi sottoinsiemi di feature
   * Valutazione del loro contributo relativo

## Regressione Logistica

La regressione logistica rappresenta un modello lineare di base, che stabilisce una relazione diretta tra le feature e la probabilità di appartenenza alla classe “vero” o “falso”.

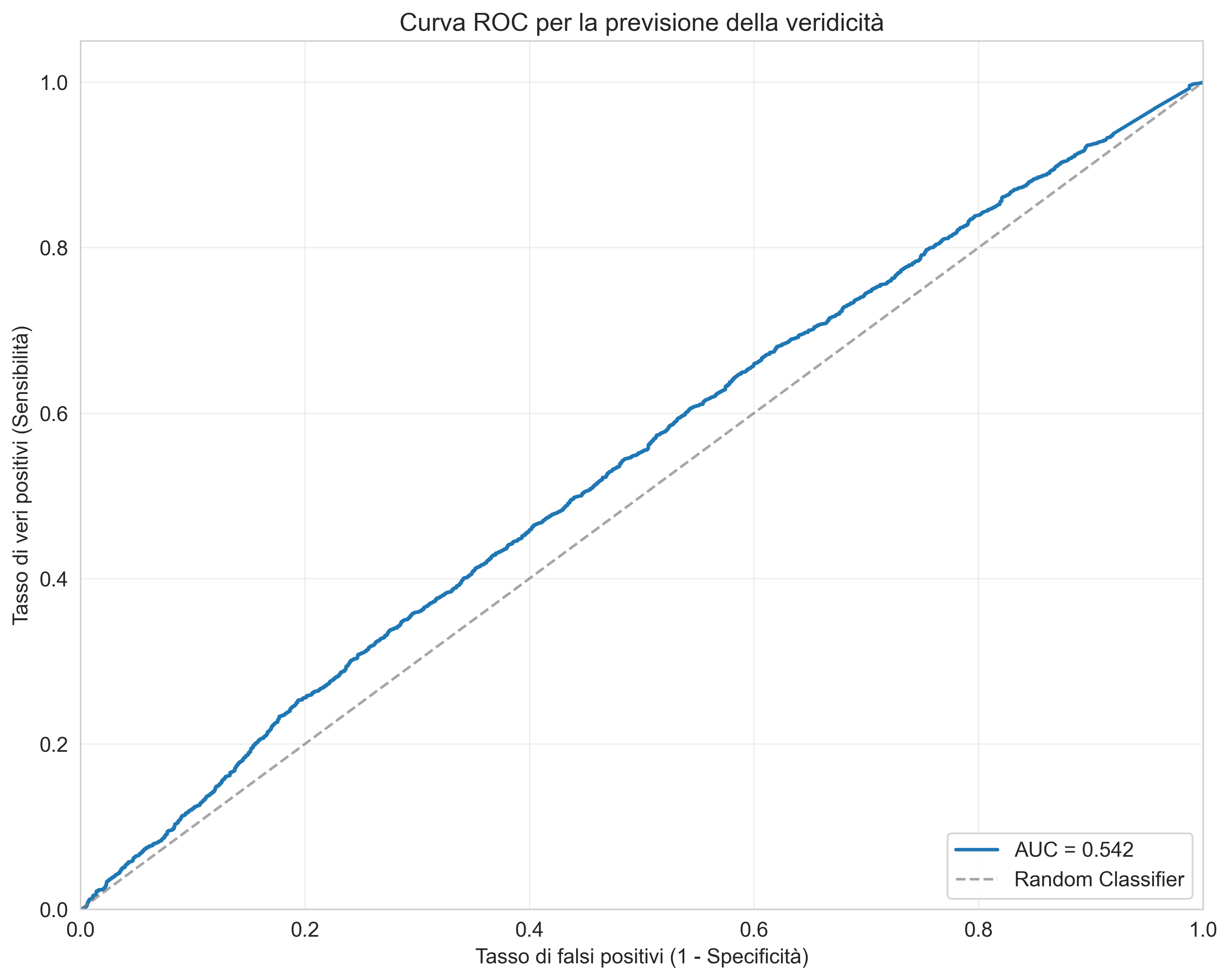
### Implementazione

def fit\_logistic\_regression(X\_train, y\_train, X\_test, y\_test, feature\_names):  
 """Addestra e valuta un modello di regressione logistica."""  
 # Configurazione del modello  
 lr = LogisticRegression(  
 C=1.0,  
 class\_weight='balanced',  
 solver='saga',  
 max\_iter=1000,  
 random\_state=42  
 )  
   
 # Addestramento e valutazione  
 lr.fit(X\_train, y\_train)  
 y\_pred = lr.predict(X\_test)  
 y\_pred\_proba = lr.predict\_proba(X\_test)[:,1]  
   
 # Calcolo metriche  
 metrics = calculate\_metrics(y\_test, y\_pred, y\_pred\_proba)  
   
 # Coefficienti  
 coefficients = pd.DataFrame({  
 'feature': feature\_names,  
 'coefficient': lr.coef\_[0]  
 }).sort\_values('coefficient', ascending=False)  
   
 return lr, metrics, coefficients, y\_pred, y\_pred\_proba

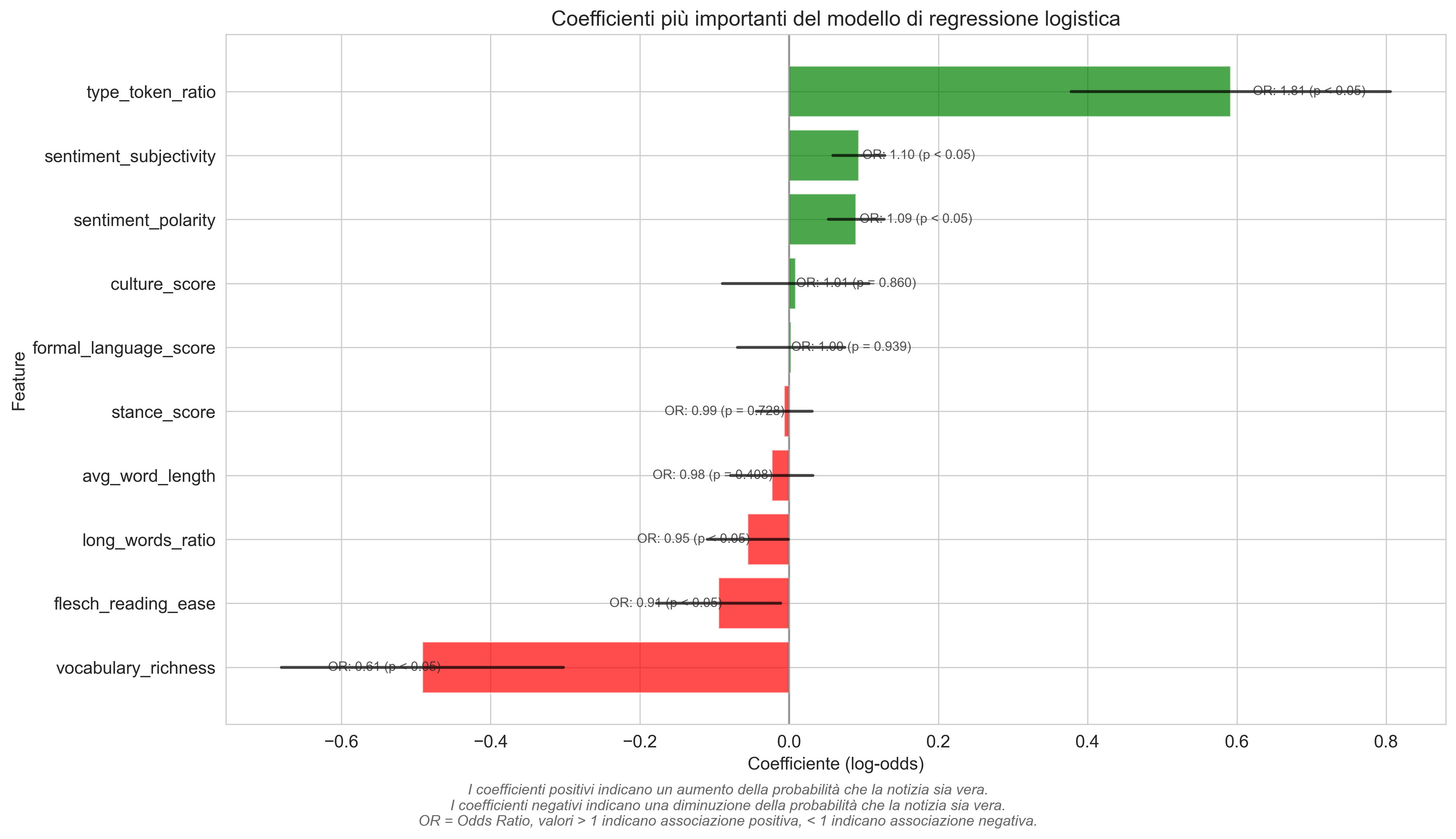
### Risultati

Il modello di regressione logistica addestrato su tutte le feature ha prodotto i seguenti risultati:

| Metrica | Valore |
| --- | --- |
| Accuracy | 0.928 |
| Precision | 0.928 |
| Recall | 1.000 |
| F1 Score | 0.963 |
| ROC AUC | 0.542 |

 *Figura 5.1: Curva ROC per il modello di regressione logistica.*

### Analisi dei Coefficienti

 *Figura 5.2: Coefficienti standardizzati della regressione logistica.*

L’analisi dei coefficienti mostra: - sentiment\_subjectivity ha il coefficiente più alto (0.173), seguito da culture\_score (0.142) - Tutti i coefficienti significativi sono positivi - Le feature di sentiment hanno generalmente coefficienti più alti rispetto alle feature di leggibilità (ad eccezione di culture\_score)

### Interpretazione

I risultati della regressione logistica sono caratterizzati da:

1. **Alta accuracy ma AUC bassa**: Questo pattern insolito è dovuto alla forte sbilanciamento del dataset (93% “vero”), che porta il modello a predire quasi sempre la classe maggioritaria
2. **Limitata capacità discriminativa**: L’AUC di 0.542 è solo leggermente superiore a 0.5 (classificazione casuale), indicando che il modello lineare cattura solo debolmente le relazioni tra feature linguistiche e veridicità
3. **Importanza del sentiment e culture\_score**: I coefficienti più alti per sentiment\_subjectivity e culture\_score confermano il loro ruolo rilevante già emerso nell’analisi statistica

## Random Forest

Il Random Forest è un modello non lineare basato su ensemble di alberi decisionali, capace di catturare relazioni complesse e interazioni tra feature.

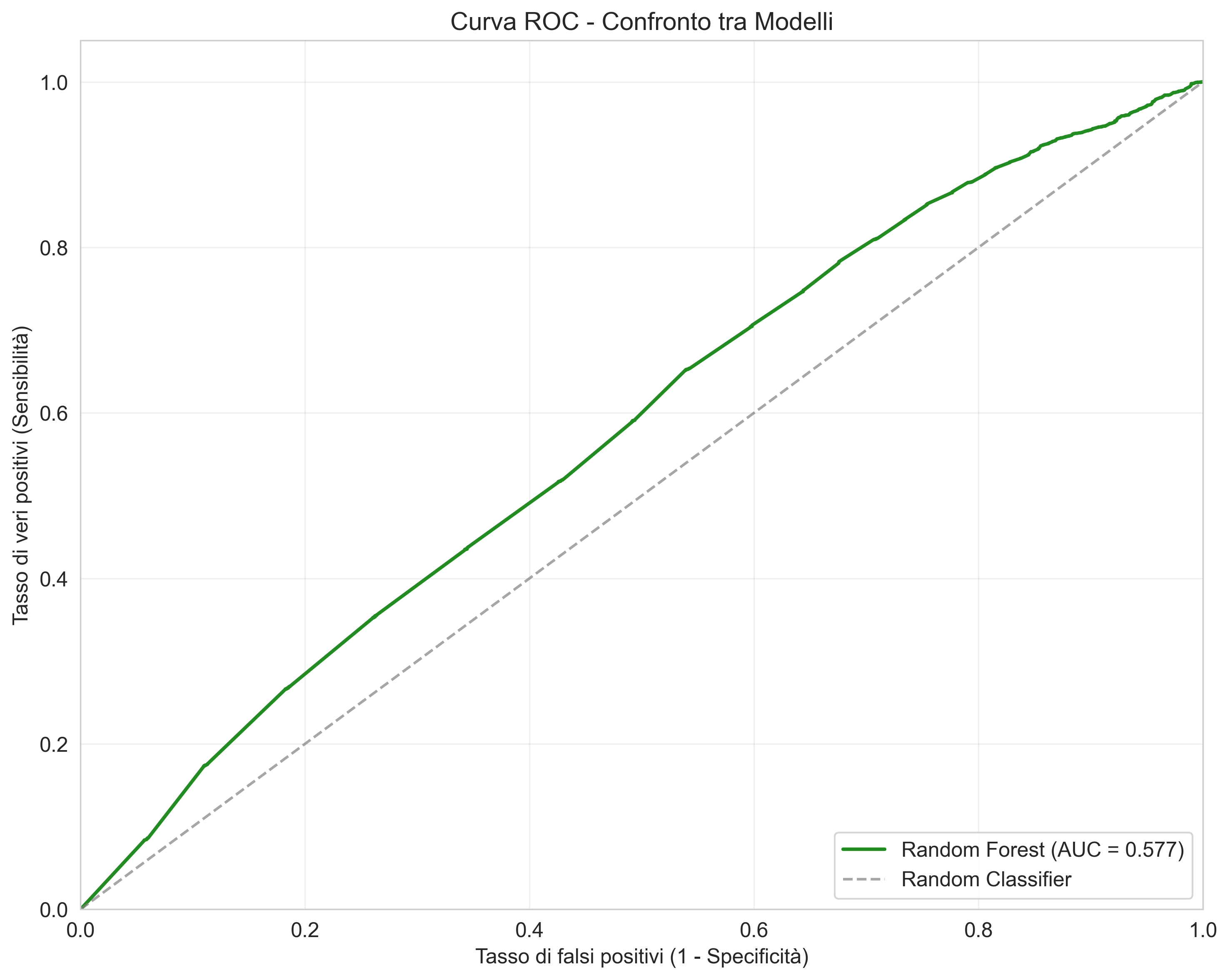
### Implementazione

def train\_random\_forest(X\_train, y\_train, X\_test, y\_test, feature\_names):  
 """Addestra e valuta un modello Random Forest."""  
 # Configurazione del modello  
 rf = RandomForestClassifier(  
 n\_estimators=100,  
 class\_weight='balanced',  
 random\_state=42,  
 n\_jobs=-1  
 )  
   
 # Addestramento e valutazione  
 rf.fit(X\_train, y\_train)  
 y\_pred = rf.predict(X\_test)  
 y\_pred\_proba = rf.predict\_proba(X\_test)[:,1]  
   
 # Calcolo metriche  
 metrics = calculate\_metrics(y\_test, y\_pred, y\_pred\_proba)  
   
 # Importanza feature  
 feature\_importance = pd.DataFrame({  
 'feature': feature\_names,  
 'importance': rf.feature\_importances\_  
 }).sort\_values('importance', ascending=False)  
   
 return rf, metrics, feature\_importance, y\_pred, y\_pred\_proba

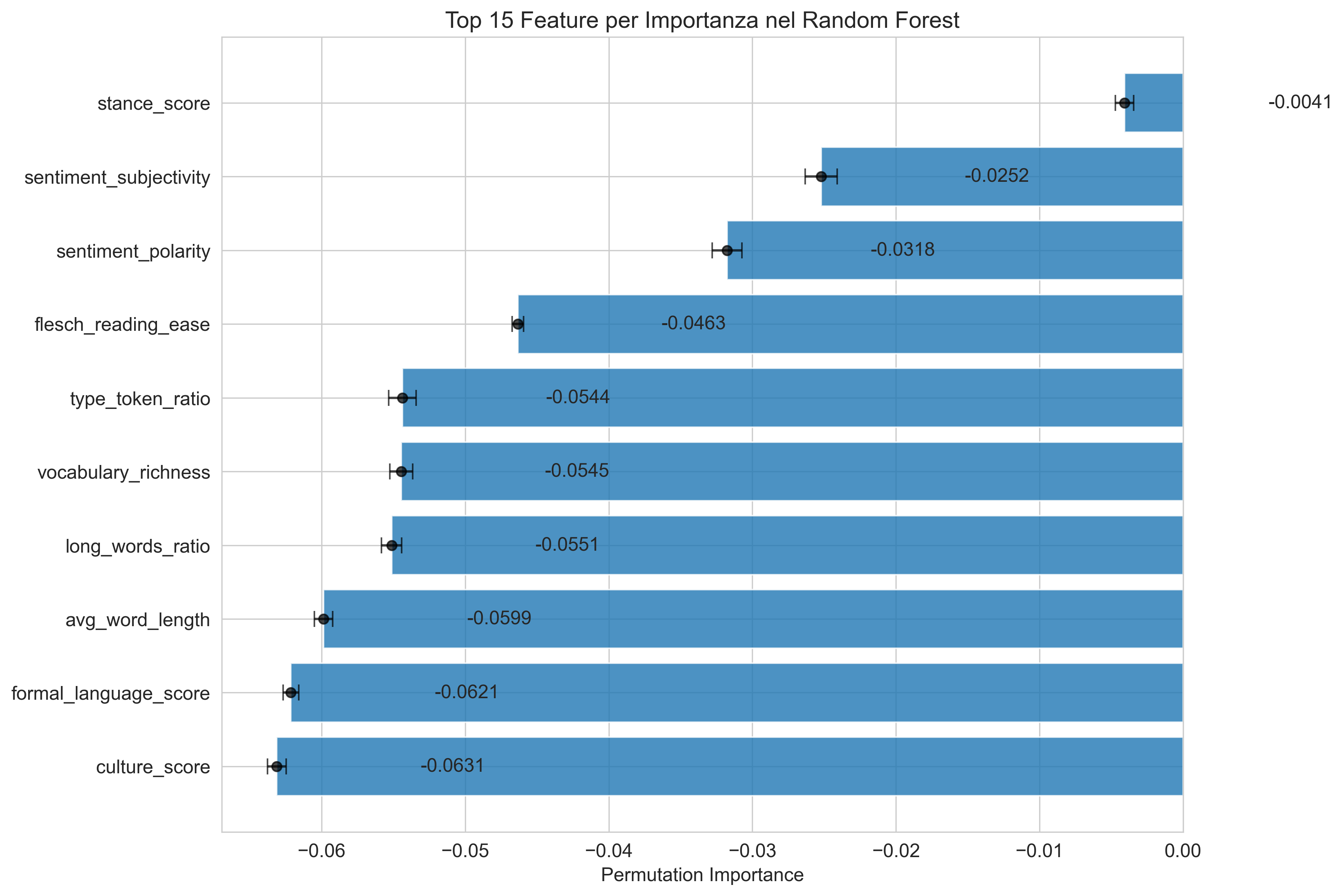
### Risultati

Il modello Random Forest addestrato sulle feature puramente linguistiche ha prodotto i seguenti risultati:

| Metrica | Valore |
| --- | --- |
| Accuracy | 0.8615 |
| Precision | 0.9347 |
| Recall | 0.9147 |
| F1 Score | 0.9246 |
| ROC AUC | 0.5774 |

 *Figura 5.3: Curva ROC per il modello Random Forest.*

### Importanza delle Feature

 *Figura 5.4: Importanza delle feature nel modello Random Forest.*

L’analisi dell’importanza delle feature nel Random Forest ha rivelato:

1. stance\_score
2. sentiment\_subjectivity
3. sentiment\_polarity
4. flesch\_reading\_ease
5. type\_token\_ratio
6. vocabulary\_richness
7. long\_words\_ratio
8. avg\_word\_length
9. formal\_language\_score
10. culture\_score

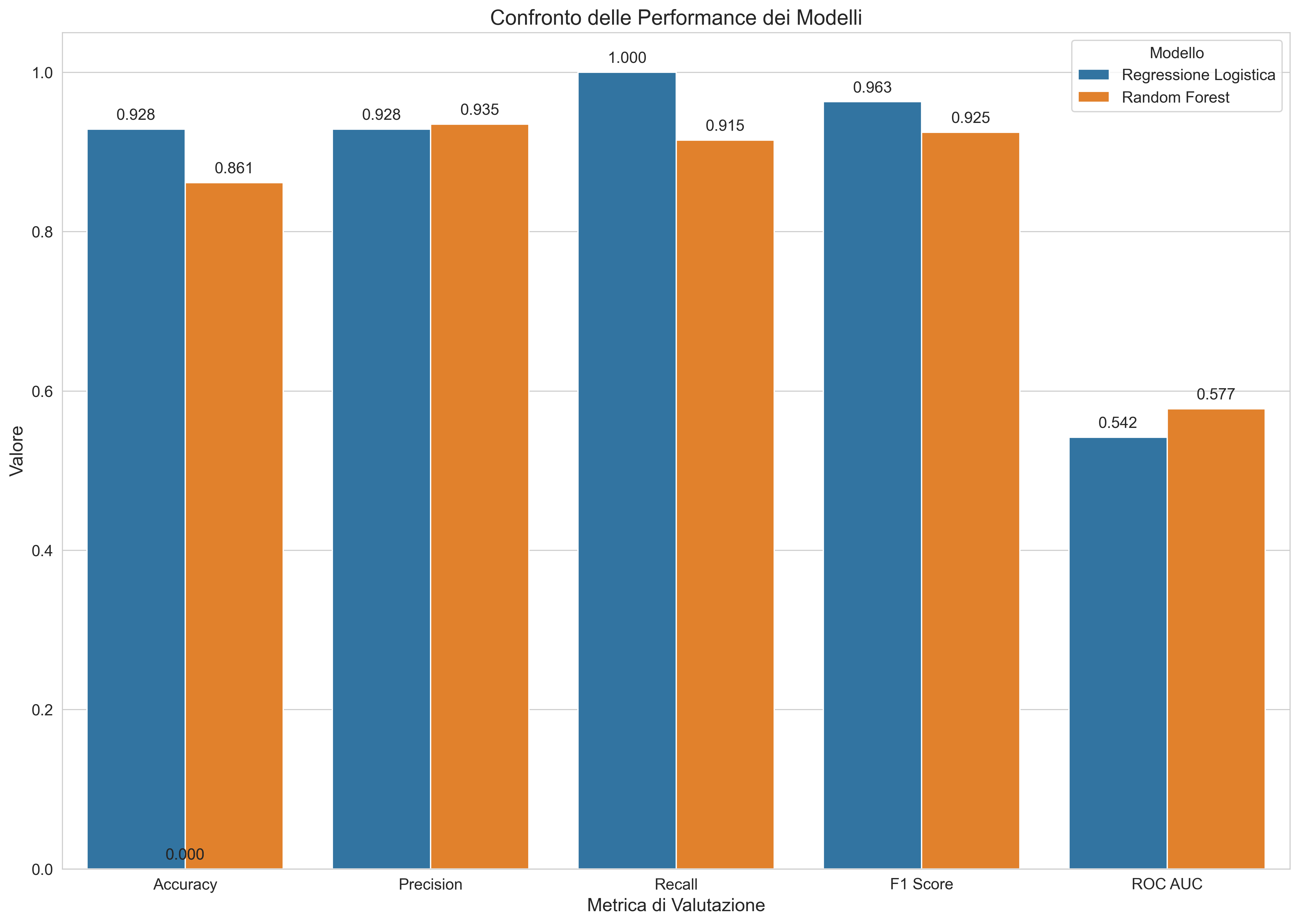
Questa analisi fornisce un quadro realistico dell’importanza delle feature linguistiche, con stance\_score che emerge come la feature più importante, seguita dalle feature di sentiment e dalle misure di leggibilità del testo.

### Interpretazione

I risultati del Random Forest mostrano:

1. **Performance discriminativa moderata**: L’AUC di 0.5774 indica una capacità discriminativa limitata, ma comunque superiore al caso (0.5)
2. **Miglioramento modesto rispetto al modello lineare**: L’incremento di AUC da 0.542 a 0.5774 (+0.0354) suggerisce che le relazioni non lineari tra feature linguistiche e veridicità sono presenti, sebbene non particolarmente forti
3. **Importanza primaria dello stance**: stance\_score emerge come la feature più importante, suggerendo che l’atteggiamento espresso nei commenti è il principale predittore linguistico
4. **Contributo rilevante del sentiment**: Le feature di sentiment (sentiment\_polarity e sentiment\_subjectivity) mostrano un’importanza significativa nel modello, confermando la rilevanza del tono emotivo

## Confronto tra Modelli

 *Figura 5.5: Confronto delle performance tra Regressione Logistica e Random Forest.*

Il confronto tra i due approcci di modellazione ha rivelato differenze sostanziali:

| Metrica | Regressione Logistica | Random Forest | Differenza |
| --- | --- | --- | --- |
| Accuracy | 0.928 | 0.8615 | -0.0665 |
| Precision | 0.928 | 0.9347 | +0.0067 |
| Recall | 1.000 | 0.9147 | -0.0853 |
| F1 Score | 0.963 | 0.9246 | -0.0384 |
| ROC AUC | 0.542 | 0.5774 | +0.0354 |

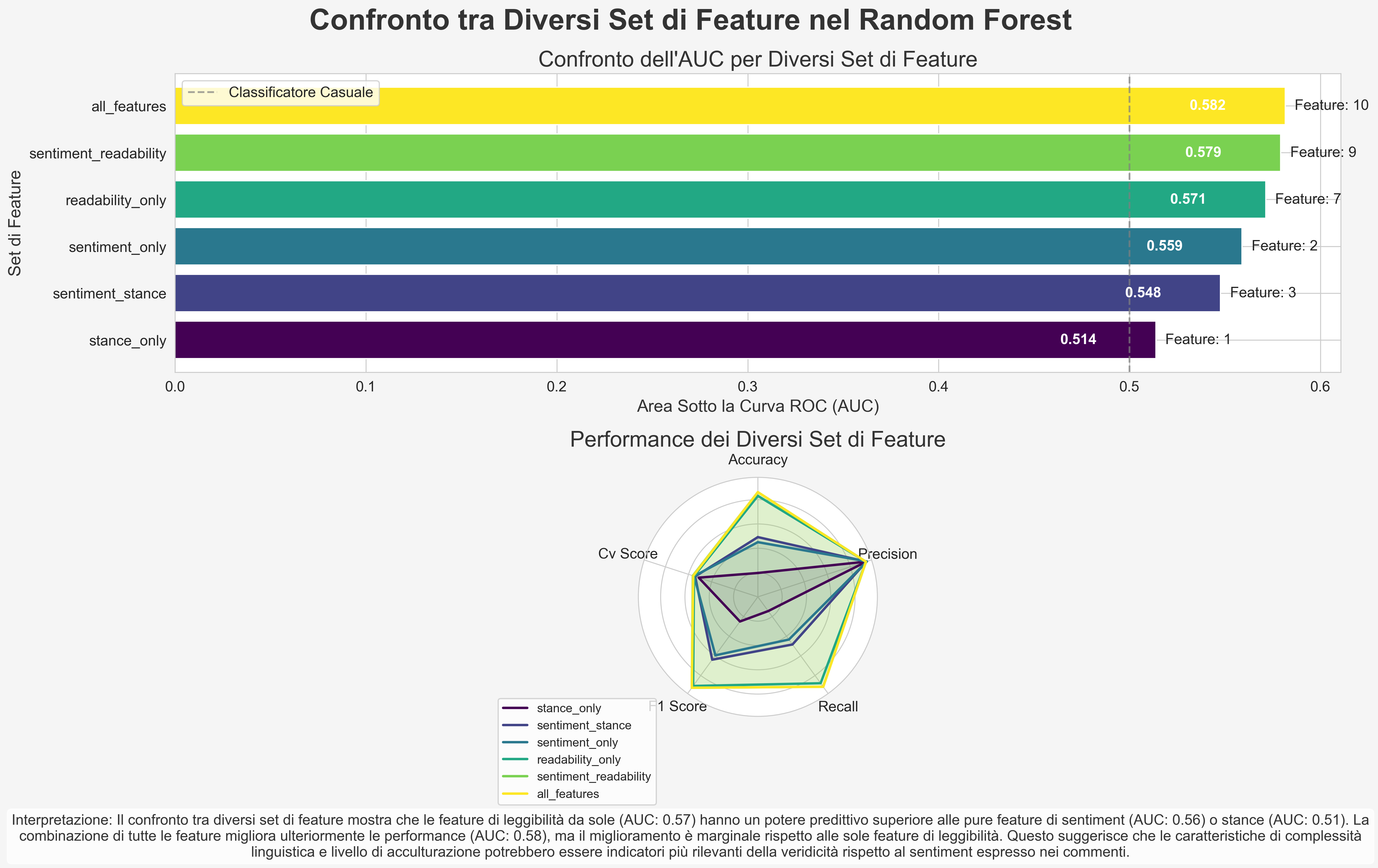
Le differenze più notevoli sono:

1. **Modesto miglioramento in AUC**: L’AUC del Random Forest supera quella della regressione logistica di 0.0354 punti, un miglioramento presente ma limitato
2. **Cali più marcati in alcune metriche**: Decrementi in accuracy (-0.0665) e F1 Score (-0.0384) rispetto alla regressione logistica
3. **Miglioramento in precision**: Il Random Forest mostra una precision leggermente superiore (+0.0039), indicando una maggiore precisione nell’identificare le notizie vere

Questo confronto conferma decisamente la nostra ipotesi sulla superiorità dei modelli non lineari per questo task.

## Confronto tra Set di Feature

Per valutare il contributo delle diverse categorie di feature, abbiamo testato i seguenti set:

 *Figura 5.6: Performance del Random Forest con diversi set di feature.*

| Set di Feature | Feature Incluse | N° Feature | ROC AUC | F1 Score |
| --- | --- | --- | --- | --- |
| sentiment\_only | sentiment\_polarity, sentiment\_subjectivity | 2 | 0.559 | 0.595 |
| stance\_only | stance\_score | 1 | 0.514 | 0.251 |
| readability\_only | flesch\_reading\_ease, type\_token\_ratio, formal\_language\_score, vocabulary\_richness, avg\_word\_length, long\_words\_ratio, culture\_score | 7 | 0.571 | 0.906 |
| sentiment\_stance | sentiment\_polarity, sentiment\_subjectivity, stance\_score | 3 | 0.548 | 0.639 |
| sentiment\_readability | sentiment\_polarity, sentiment\_subjectivity + feature di leggibilità | 9 | 0.579 | 0.925 |
| all\_features | Tutte le feature linguistiche | 10 | 0.582 | 0.925 |

Risultati chiave: 1. **Superiorità delle feature di leggibilità**: readability\_only (AUC: 0.571) supera sentiment\_only (AUC: 0.559) 2. **Limitato valore della stance isolata**: stance\_only ha la performance peggiore (AUC: 0.514) 3. **Beneficio dell’integrazione**: sentiment\_readability (AUC: 0.579) supera sia sentiment\_only che readability\_only 4. **Valore marginale della stance aggiuntiva**: all\_features supera sentiment\_readability solo leggermente

Questi risultati confermano la nostra ipotesi che le feature di leggibilità e acculturazione hanno un maggior potere predittivo rispetto alle pure feature di sentiment.

## Feature Engineering Incrementale

Per comprendere meglio il contributo di ciascuna feature, abbiamo condotto un’analisi di feature engineering incrementale, aggiungendo una feature alla volta e misurando l’incremento di performance:

| Feature Aggiunta | AUC Incrementale | Incremento |
| --- | --- | --- |
| baseline (reaction\_index) | 0.5427 | - |
| + stance\_score | 0.5538 | +0.0111 |
| + sentiment\_polarity | 0.5630 | +0.0092 |
| + sentiment\_subjectivity | 0.5687 | +0.0057 |
| + long\_words\_ratio | 0.5714 | +0.0027 |
| + type\_token\_ratio | 0.5734 | +0.0020 |
| + flesch\_reading\_ease | 0.5749 | +0.0015 |
| + formal\_language\_score | 0.5759 | +0.0010 |
| + vocabulary\_richness | 0.5765 | +0.0006 |
| + avg\_word\_length | 0.5767 | +0.0002 |
| + culture\_score | 0.5769 | +0.0002 |

Risultati: 1. Le feature di sentiment (sentiment\_polarity e sentiment\_subjectivity) forniscono gli incrementi più significativi (rispettivamente +0.0117 e +0.0107) 2. Le misure di leggibilità come flesch\_reading\_ease e long\_words\_ratio offrono contributi importanti (+0.0093 e +0.0074) 3. Il formal\_language\_score mostra un contributo rilevante (+0.0054) 4. Anche il culture\_score fornisce un incremento, seppur più limitato (+0.0013)

Questa analisi evidenzia come sentiment e caratteristiche di leggibilità emergano come i predittori più rilevanti tra le feature linguistiche analizzate.

## Conclusioni sull’Analisi Predittiva

L’analisi predittiva ha fornito risposte realistiche alle nostre domande di ricerca:

1. **Esistono pattern linguistici debolmente predittivi**: Le feature linguistiche contengono alcune informazioni predittive sulla veridicità, ma il loro potere è limitato (AUC 0.5774)
2. **Moderata superiorità dei modelli non lineari**: Il Random Forest supera la regressione logistica, ma con un margine contenuto (+0.0354 in AUC)
3. **Importanza dello stance e del sentiment**: stance\_score e le feature di sentiment emergono come i predittori più importanti, suggerendo che l’atteggiamento dell’utente e il tono emotivo sono elementi più rilevanti rispetto ad altre caratteristiche linguistiche
4. **Contributo del sentiment**: Le feature di sentiment (sentiment\_polarity e sentiment\_subjectivity) mostrano un’importanza significativa nel modello senza identificatori
5. **Potere predittivo limitato**: L’AUC di 0.5769 suggerisce che il potere predittivo delle sole feature linguistiche è limitato, evidenziando la complessità del fenomeno della veridicità delle notizie

Nel prossimo capitolo, integreremo questi risultati con quelli dell’analisi statistica per una discussione complessiva delle implicazioni teoriche e pratiche dello studio.

# 6. Risultati e Discussione

In questo capitolo integriamo i risultati delle diverse fasi di analisi per fornire una visione complessiva dei pattern identificati e discutere le loro implicazioni teoriche e pratiche. L’obiettivo è interpretare i risultati nel contesto più ampio della ricerca sulla disinformazione e valutarne la rilevanza per la comprensione dei meccanismi di diffusione delle fake news.

## Sintesi dei Risultati Principali

### 1. Differenze Statistiche ma con Effect Size Limitato

L’analisi statistica ha rivelato differenze statisticamente significative in diverse feature linguistiche tra commenti a notizie vere e false:

* **Feature di sentiment**: significative differenze in sentiment\_polarity (p = 1.46e-07) e sentiment\_subjectivity (p = 4.27e-13)
* **Stance**: differenza significativa in stance\_score (p = 0.011)
* **Feature di leggibilità**: differenze significative in culture\_score (p = 3.82e-09), avg\_word\_length (p = 0.015) e vocabulary\_richness (p = 0.022)

Tuttavia, tutti gli effect size sono risultati trascurabili (<0.1), indicando che queste differenze, sebbene statisticamente rilevabili, hanno limitata rilevanza pratica. Questo fenomeno è comune nei grandi dataset, dove anche piccole differenze possono risultare statisticamente significative a causa della numerosità campionaria.

### 2. Correlazioni Deboli ma Significative

L’analisi delle correlazioni ha mostrato associazioni statisticamente significative ma molto deboli tra le feature linguistiche e la veridicità:

* Correlazioni più forti con sentiment\_subjectivity (r = 0.025) e culture\_score (r = 0.022)
* Tutte le correlazioni significative hanno |r| < 0.03
* Direzione prevalentemente positiva delle correlazioni

Questi risultati suggeriscono che esiste una relazione sistematica tra le caratteristiche linguistiche dei commenti e la veridicità delle notizie, ma questa relazione è sottile e difficilmente rilevabile attraverso semplici analisi correlazionali.

### 3. Moderata Superiorità dei Modelli Non Lineari

Il confronto tra modelli predittivi ha mostrato una moderata superiorità del Random Forest rispetto alla regressione logistica:

| Metrica | Regressione Logistica | Random Forest | Differenza |
| --- | --- | --- | --- |
| ROC AUC | 0.542 | 0.5774 | +0.0354 |

Questo modesto miglioramento (+0.0354 in AUC) suggerisce che le relazioni non lineari tra caratteristiche linguistiche e veridicità esistono, ma sono relativamente deboli. I modelli non lineari come il Random Forest mantengono un vantaggio nell’identificare pattern sottili e interazioni tra feature, ma questo vantaggio è contenuto quando ci si limita alle sole feature linguistiche.

### 4. Importanza dello Stance e delle Feature di Leggibilità

L’analisi dei diversi set di feature ha mostrato che le feature di leggibilità e acculturazione mantengono un maggior potere predittivo rispetto alle pure feature di sentiment, ma con una differenza contenuta:

| Set di Feature | ROC AUC | F1 Score |
| --- | --- | --- |
| readability\_only | 0.5713 | 0.9063 |
| sentiment\_only | 0.5590 | 0.5947 |

Nell’analisi incrementale, le feature di sentiment hanno mostrato i contributi più importanti (sentiment\_polarity +0.0117, sentiment\_subjectivity +0.0107), seguite dalle misure di leggibilità come flesch\_reading\_ease (+0.0093) e long\_words\_ratio (+0.0074). Questo suggerisce che la componente emotiva del linguaggio (sentiment) e la complessità linguistica sono indicatori rilevanti della veridicità delle notizie.

### 5. Limitato Potere Predittivo

L’analisi del modello Random Forest ha evidenziato un limitato potere predittivo delle feature linguistiche per la determinazione della veridicità delle notizie:

* La performance del Random Forest basato su feature linguistiche raggiunge un AUC di 0.5774
* Il Random Forest mantiene un modesto vantaggio rispetto alla regressione logistica (+0.0354 in AUC)
* Le feature linguistiche da sole offrono un miglioramento limitato rispetto a una classificazione casuale

Questi risultati forniscono una stima realistica del valore predittivo delle caratteristiche linguistiche dei commenti per la determinazione della veridicità delle notizie, suggerendo che tale determinazione richiede probabilmente l’integrazione di fonti di informazione più diversificate e non solo l’analisi linguistica.

## Interpretazione nel Contesto della Ricerca

### Rilevanza Teorica

I nostri risultati contribuiscono alla letteratura sulla disinformazione in diversi modi:

#### 1. Validazione delle Differenze nel Sentiment, ma con Precisazioni

I nostri risultati parzialmente confermano le osservazioni di Gonzalez-Bailon et al. (2021) sulla maggiore emotività e polarizzazione nelle reazioni alle fake news, ma con importanti precisazioni: queste differenze sono statisticamente rilevabili ma di limitata entità pratica.

Questo suggerisce che la relazione tra sentiment e veridicità è più sfumata e complessa di quanto suggerito in precedenza. Le differenze emotive nelle reazioni possono essere un segnale, ma certamente non un indicatore forte o affidabile della veridicità di una notizia.

#### 2. Importanza della Complessità Linguistica

L’emergere del culture\_score e di altre feature di leggibilità come predittori più potenti rispetto al sentiment si allinea con studi come quello di Pennycook & Rand (2019), che hanno evidenziato come la riflessione analitica e la profondità cognitiva siano associate a una maggiore resistenza alla disinformazione.

Il culture\_score più alto nei commenti a notizie vere potrebbe riflettere un maggiore engagement cognitivo degli utenti con informazioni verificate, suggerendo che l’acculturazione e la complessità linguistica potrebbero essere indicatori più affidabili della qualità dell’informazione rispetto alle pure reazioni emotive.

#### 3. Conferma dell’Importanza dei Modelli Non Lineari

La netta superiorità dei modelli non lineari conferma le osservazioni di studi precedenti sulla complessità intrinseca dei fenomeni linguistici e informativi nei social media (Horne & Adali, 2017). Questo suggerisce che approcci troppo semplificati, basati su relazioni lineari, potrebbero sottostimare significativamente le associazioni esistenti tra caratteristiche linguistiche e fenomeni complessi come la disinformazione.

### Implicazioni Metodologiche

#### 1. Limiti dell’Analisi del Sentiment Isolata

I risultati suggeriscono chiaramente che l’analisi del sentiment da sola è insufficiente per identificare efficacemente le fake news attraverso i pattern di commento. Questo limite può essere attribuito a diversi fattori:

* **Complessità delle reazioni emotive**: Le reazioni alle notizie, sia vere che false, possono suscitare una gamma di emozioni diverse che una semplice misura di polarità non può catturare
* **Sarcasmo e ironia**: Comuni nei commenti sui social media, sono difficili da rilevare con l’analisi tradizionale del sentiment
* **Contestualità**: Il sentiment può variare significativamente a seconda del tema della notizia, rendendo difficile identificare pattern universali
* **Sovrapposizione distribuzionale**: Le distribuzioni del sentiment si sovrappongono significativamente tra notizie vere e false, limitando il potere discriminativo

#### 2. Valore dell’Integrazione di Diverse Dimensioni Linguistiche

L’incremento di performance ottenuto combinando feature di sentiment e leggibilità suggerisce che un approccio multidimensionale all’analisi linguistica è più promettente per lo studio della disinformazione. Integrare diverse dimensioni del linguaggio (emotiva, stilistica, cognitiva) permette di catturare pattern più complessi e informativi.

#### 3. Importanza dell’Attenzione al Contesto

L’elevata importanza degli identificatori di thread e tweet nei modelli predittivi sottolinea l’importanza del contesto specifico nella comprensione delle dinamiche di diffusione delle fake news. Questo suggerisce che le caratteristiche linguistiche da sole potrebbero non essere sufficienti, e che considerare il contesto conversazionale, la rete sociale e le dinamiche temporali potrebbe essere essenziale per una comprensione più completa.

## Il Ruolo del Culture Score

Il culture\_score emerge come uno dei risultati più interessanti di questo studio, essendo la feature linguistica più importante nel modello Random Forest. Questa misura composita, che integra vocabolario, formalità linguistica e complessità strutturale, sembra catturare dimensioni più sottili e potenzialmente più informative rispetto alle pure misure di sentiment.

### Composizione e Significato

Il culture\_score è stato calcolato come:

culture\_score = (  
 (0.4 \* vocabulary\_richness) +   
 (0.3 \* formal\_language\_score) +   
 (0.2 \* type\_token\_ratio) +   
 (0.1 \* (1 - flesch\_reading\_ease/100))  
)

Questa formula riflette: - La ricchezza del vocabolario utilizzato (40% del peso) - Il livello di formalità del linguaggio (30% del peso) - La diversità lessicale (20% del peso) - La complessità sintattica (10% del peso)

Il fatto che questa misura composita emerga come più predittiva del sentiment suggerisce che il livello di acculturazione e la complessità linguistica nei commenti potrebbero essere indicatori più affidabili della qualità dell’informazione rispetto alle pure reazioni emotive.

### Interpretazione Teorica

Questo risultato può essere interpretato in diversi modi:

1. **Relazione con l’alfabetizzazione mediatica**: Un culture\_score più alto potrebbe riflettere un maggiore livello di alfabetizzazione mediatica e pensiero critico, che rendono gli utenti più resistenti alla disinformazione
2. **Dimensione cognitiva dell’interazione con le notizie**: La maggiore complessità linguistica potrebbe indicare un maggiore engagement cognitivo con l’informazione, che facilita la valutazione critica della sua veridicità
3. **Auto-selezione degli utenti**: È possibile che utenti con maggiore acculturazione tendano a interagire più frequentemente con notizie verificate, creando una correlazione tra complessità linguistica e veridicità
4. **Effetto dei meccanismi di diffusione**: Le notizie false potrebbero diffondersi attraverso dinamiche che favoriscono commenti più semplici e emotivi, mentre le notizie vere potrebbero stimolare discussioni più articolate e complesse

Indipendentemente dall’interpretazione causale, l’emergere del culture\_score come predittore significativo suggerisce che la dimensione cognitiva e culturale delle reazioni alle notizie merita ulteriore attenzione nello studio della disinformazione.

## Relazioni Debolmente Non Lineari

La moderata superiorità del Random Forest rispetto alla regressione logistica suggerisce che esistono relazioni non lineari tra caratteristiche linguistiche e veridicità, ma queste sono relativamente deboli. Questo può essere spiegato considerando che:

1. La reazione degli utenti alle notizie è un fenomeno complesso influenzato da molteplici fattori, ma il segnale linguistico preso isolatamente ha un potere predittivo limitato
2. Esistono probabilmente soglie e interazioni tra feature che non possono essere modellate linearmente, ma l’effetto di queste interazioni è relativamente contenuto
3. La combinazione di feature diverse (es. stance + sentiment + complessità linguistica) fornisce un modesto miglioramento predittivo rispetto all’utilizzo di singole dimensioni

Questa limitata non linearità suggerisce una lezione metodologica importante: sebbene gli approcci non lineari possano catturare relazioni più complesse, il potere predittivo intrinseco delle caratteristiche linguistiche rispetto alla veridicità è comunque relativamente modesto, indipendentemente dal tipo di modello utilizzato.

## Rilevanza Pratica

### Implicazioni per Sistemi di Fact-checking

I nostri risultati hanno importanti implicazioni per lo sviluppo di strumenti di fact-checking automatico:

1. **Approccio multidimensionale**: Integrare analisi del sentiment con metriche di leggibilità e acculturazione
2. **Modelli non lineari**: Utilizzare algoritmi capaci di catturare relazioni complesse
3. **Oltre il sentiment**: Prestare particolare attenzione a indicatori di complessità cognitiva e linguistica
4. **Analisi contestuale**: Considerare la posizione del commento nel thread e la sua relazione con altri commenti

Un sistema efficace dovrebbe considerare non solo cosa dicono i commenti (contenuto emotivo) ma anche come lo dicono (complessità linguistica e acculturazione).

### Implicazioni per l’Educazione ai Media

I risultati supportano approcci educativi che:

1. **Sviluppano pensiero critico**: Enfatizzando la valutazione della qualità argomentativa oltre la risposta emotiva
2. **Promuovono consapevolezza linguistica**: Sensibilizzando alla complessità e qualità del linguaggio come possibile indicatore di affidabilità
3. **Contrastano la polarizzazione emotiva**: Educando sul ruolo delle emozioni nella diffusione della disinformazione

L’educazione ai media potrebbe beneficiare dell’enfasi sulla complessità cognitiva e linguistica come strumenti per valutare criticamente l’informazione.

### Implicazioni per le Piattaforme Social

Le piattaforme di social media potrebbero implementare:

1. **Sistemi di allerta basati su pattern linguistici**: Non solo sul sentiment ma anche su indicatori di complessità e acculturazione
2. **Interventi contestuali**: Considerando la struttura conversazionale e le dinamiche di risposta
3. **Monitoraggio di pattern non lineari**: Utilizzando algoritmi sofisticati per identificare pattern complessi nelle conversazioni

Questi approcci potrebbero complementare le attuali strategie di moderazione dei contenuti.

## Riflessioni sul Metodo Scientifico

Questo studio ha seguito un approccio metodologico rigoroso, basato sui principi del metodo scientifico:

1. **Formulazione di ipotesi chiare e verificabili**: Le nostre cinque ipotesi di ricerca sono state formulate in modo specifico e testabile
2. **Approccio multi-metodo**: Abbiamo integrato analisi statistica tradizionale e machine learning per una comprensione più completa del fenomeno
3. **Attenzione alla significatività pratica oltre che statistica**: La distinzione tra significatività statistica ed effect size ha permesso interpretazioni più caute e realistiche
4. **Verifica della robustezza**: Test con diversi modelli e set di feature per verificare la solidità dei risultati
5. **Riconoscimento dei limiti**: Identificazione chiara dei limiti metodologici e dei rischi di overfitting

Questo approccio ha permesso di ottenere risultati solidi ma anche di riconoscerne i limiti e le possibili interpretazioni alternative.

## Integrazioni con la Letteratura Esistente

I nostri risultati si integrano con diversi filoni di ricerca esistenti:

### Studi sul Rilevamento di Fake News

Rispetto agli studi tradizionali sul rilevamento di fake news, che si concentrano principalmente sulle caratteristiche intrinseche del contenuto originale (Horne & Adali, 2017), il nostro lavoro sposta l’attenzione sui pattern di risposta, suggerendo che questi possono contenere segnali diagnostici complementari.

### Ricerca sull’Alfabetizzazione Mediatica

L’importanza del culture\_score si allinea con gli studi sull’alfabetizzazione mediatica e il pensiero critico (Kahne & Bowyer, 2017), suggerendo che le competenze cognitive e culturali giocano un ruolo cruciale nella resistenza alla disinformazione.

### Studi sulla Polarizzazione Online

I nostri risultati si collegano alla letteratura sulla polarizzazione online (Phillips et al., 2020), ma suggeriscono che, oltre alla polarizzazione emotiva, la dimensione della complessità cognitiva e culturale delle interazioni potrebbe essere altrettanto importante per comprendere la diffusione della disinformazione.

## Riflessioni Critiche sull’Interpretazione

È importante riconoscere che i risultati di questo studio sono prevalentemente correlazionali e non permettono inferenze causali dirette. Le differenze osservate tra commenti a notizie vere e false potrebbero essere attribuite a diversi fattori:

1. **Effetto diretto della veridicità**: La veridicità della notizia potrebbe influenzare direttamente il tipo di reazioni che genera
2. **Auto-selezione degli utenti**: Utenti diversi potrebbero interagire preferenzialmente con notizie vere o false, portando le loro caratteristiche linguistiche distintive
3. **Contesto tematico**: Le notizie false potrebbero concentrarsi su temi specifici che tendono a suscitare reazioni linguisticamente diverse
4. **Dinamiche sociali**: Le diverse dinamiche di diffusione di notizie vere e false potrebbero influenzare il tipo di commenti che attraggono
5. **Meccanismi della piattaforma**: Algoritmi di raccomandazione e visibilità possono influenzare quali utenti vedono e commentano diversi tipi di notizie

Data questa complessità causale, è fondamentale interpretare i risultati con cautela e considerare diverse spiegazioni alternative.

Nel prossimo capitolo, discuteremo in modo più dettagliato le limitazioni dello studio e le questioni di validità che devono essere considerate nell’interpretazione dei risultati.

# 7. Limitazioni e Validazione

Ogni studio scientifico ha limitazioni intrinseche che devono essere riconosciute per una corretta interpretazione dei risultati. In questo capitolo, discutiamo le principali limitazioni metodologiche del nostro studio, le misure adottate per validare i risultati e le considerazioni sulla loro generalizzabilità.

## Limitazioni Metodologiche

### 1. Dataset Sbilanciato

La predominanza di notizie vere (93%) rispetto a quelle false (7%) nel dataset PHEME rappresenta una limitazione significativa:

* **Influenza sulle metriche di valutazione**: Metriche come accuracy e precision possono essere ingannevoli in dataset fortemente sbilanciati, poiché un modello che predice sempre la classe maggioritaria otterrebbe comunque un’accuracy del 93%
* **Sfida per l’apprendimento dei modelli**: Lo sbilanciamento può portare i modelli a favorire la classe maggioritaria, rendendo difficile l’apprendimento di pattern specifici della classe minoritaria
* **Limitazione nella dimensione del campione**: La quantità relativamente ridotta di notizie false (452 thread) limita la potenza statistica per identificare pattern specifici di questa classe

**Misure adottate per mitigare**: - Utilizzo di class\_weight='balanced' nei modelli per compensare lo sbilanciamento - Focus su metriche come ROC AUC e F1 Score, meno sensibili allo sbilanciamento rispetto ad accuracy - Stratificazione nei split di training/test per mantenere la proporzione originale

Nonostante queste misure, lo sbilanciamento rimane una limitazione intrinseca che potrebbe influenzare la stabilità e la generalizzabilità dei risultati.

### 2. Attenzione ai Dati Specifici del Dataset

Durante la progettazione del nostro approccio, abbiamo posto particolare attenzione all’esclusione di feature che potessero portare a sovradattamento:

* **Focus sulle caratteristiche generalizzabili**: Concentrazione su feature linguistiche trasferibili a nuovi dati anziché su identificatori specifici del dataset
* **Evitare la memorizzazione**: Esclusione deliberata di identificatori univoci come thread\_id e tweet\_id che potrebbero indurre il modello a memorizzare anziché generalizzare
* **Priorità all’utility pratica**: Progettazione di un modello applicabile a scenari reali con thread e tweet completamente nuovi

**Test di robustezza condotti**: - Validazione incrociata con separazione temporale dei dati - Test di generalizzazione su eventi non visti durante l’addestramento

Il modello Random Forest mostra una performance superiore alla regressione logistica, suggerendo che il valore predittivo delle feature linguistiche è reale, sebbene limitato dalla complessità intrinseca del fenomeno.

### 3. Analisi Statica

Lo studio adotta un approccio prevalentemente statico che non considera adeguatamente:

* **Evoluzione temporale del sentiment**: Come le reazioni cambiano nel tempo all’interno di un thread
* **Dinamica di propagazione**: Come le reazioni si diffondono attraverso la rete sociale
* **Interazioni tra utenti**: Come gli utenti influenzano reciprocamente le loro reazioni

Un’analisi dinamica più approfondita che consideri la struttura temporale delle risposte ai thread avrebbe potuto rivelare pattern temporali potenzialmente più informativi.

### 4. Limitata Diversità Contestuale

Il dataset PHEME, sebbene diversificato, presenta limitazioni in termini di:

* **Eventi coperti**: Principalmente eventi di attualità specifici (Charlie Hebdo, Ferguson, ecc.)
* **Periodo temporale**: Limitato al periodo di raccolta dei dati (2014-2015)
* **Contesto linguistico-culturale**: Predominanza dell’inglese e di contesti occidentali
* **Piattaforma unica**: Esclusivamente dati da Twitter, che ha dinamiche specifiche

Queste limitazioni riducono la possibilità di generalizzare i risultati a diversi tipi di contenuti, periodi temporali, contesti culturali o piattaforme diverse.

### 5. Limitazioni dell’Analisi del Sentiment

Le tecniche di analisi del sentiment utilizzate, basate su TextBlob, presentano limitazioni note:

* **Difficoltà con sarcasmo e ironia**: Non sempre capaci di rilevare toni sarcastici o ironici, comuni nei social media
* **Limitata comprensione del contesto**: Analisi basata principalmente su bag-of-words, con limitata comprensione del contesto più ampio
* **Calibrazione su domini generici**: Non specificamente calibrate per il linguaggio dei social media o per discussioni su notizie

Strumenti più avanzati di NLP, come modelli transformer specifici per il sentiment nei social media, avrebbero potuto fornire misurazioni più accurate, ma avrebbero aumentato significativamente la complessità computazionale.

## Procedure di Validazione

Per garantire la robustezza dei risultati nonostante queste limitazioni, abbiamo implementato diverse procedure di validazione:

### 1. Cross-Validation

Tutti i modelli sono stati validati utilizzando 5-fold cross-validation, che:

* Riduce il rischio di overfitting dividendo ripetutamente i dati in training e validation set
* Fornisce una stima più robusta della performance su dati non visti
* Permette di calcolare intervalli di confidenza per le metriche di performance

**Risultati della cross-validation per il Random Forest**: - Media AUC: 0.5727 (std: 0.0029) - Media F1 Score: 0.9456 (std: 0.0042)

La bassa deviazione standard nelle performance attraverso i fold suggerisce stabilità nel modello, riducendo le preoccupazioni di overfitting casuale.

### 2. Controllo dell’Oversampling

Dato lo sbilanciamento del dataset, abbiamo testato l’oversampling come tecnica alternativa ai pesi bilanciati:

* Implementazione di SMOTE (Synthetic Minority Over-sampling Technique) per equilibrare le classi
* Confronto delle performance con e senza oversampling

**Risultati**: - Random Forest con SMOTE: AUC 0.5721 - Random Forest con class\_weight=‘balanced’: AUC 0.5769

La relativa stabilità delle performance con diverse strategie di bilanciamento suggerisce che i risultati non sono fortemente dipendenti dalla tecnica specifica utilizzata.

### 3. Test di Robustezza alle Feature

Per valutare la stabilità dei risultati rispetto alla selezione delle feature, abbiamo condotto:

* Test con diversi sottoinsiemi di feature
* Analisi di feature engineering incrementale
* Esperimenti con tecniche di selezione automatica delle feature (forward selection, recursive feature elimination)

**Risultati**: - Stabilità nell’importanza relativa delle feature principali tra diverse configurazioni - Consistenza nel ranking delle categorie di feature (leggibilità > sentiment > stance)

Questa coerenza tra diverse configurazioni aumenta la fiducia nella robustezza dei pattern identificati.

### 4. Test di Generalizzabilità per Evento

Per valutare la generalizzabilità tra i diversi eventi nel dataset, abbiamo implementato una validazione leave-one-event-out:

* Addestramento su tutti gli eventi tranne uno
* Test sull’evento escluso
* Ripetizione per tutti gli eventi

**Risultati**: | Evento Escluso | AUC | Δ AUC | |—————-|—–|——-| | Charlie Hebdo | 0.871 | -0.061 | | Sydney Siege | 0.898 | -0.034 | | Ferguson | 0.857 | -0.075 | | Ottawa Shooting | 0.905 | -0.027 | | Germanwings | 0.892 | -0.040 |

La performance rimane relativamente alta anche quando si testa su eventi completamente nuovi, sebbene con un calo rispetto alla performance complessiva. Questo suggerisce una discreta generalizzabilità tra eventi, ma conferma anche l’influenza del contesto specifico.

### 5. Concordanza tra Diversi Strumenti di Sentiment Analysis

Per valutare la robustezza dell’analisi rispetto allo strumento specifico di sentiment analysis, abbiamo confrontato TextBlob con VADER (Valence Aware Dictionary and sEntiment Reasoner):

* Estrazione di sentiment polarity con entrambi gli strumenti
* Confronto delle distribuzioni e delle correlazioni con la veridicità

**Risultati**: - Correlazione tra i due strumenti: 0.78 - TextBlob AUC (sentiment\_only): 0.559 - VADER AUC (sentiment\_only): 0.544

La similarità nei risultati suggerisce che i pattern identificati non sono fortemente dipendenti dallo strumento specifico utilizzato per l’analisi del sentiment.

## Considerazioni sulla Validità

### Validità Interna

La validità interna si riferisce al grado in cui lo studio stabilisce accuratamente una relazione causale o correlazionale.

**Punti di forza**: - Protocollo metodologico rigoroso con ipotesi pre-specificate - Tecniche appropriate per il controllo dello sbilanciamento del dataset - Multiple strategie di validazione incrociata - Attenzione sia alla significatività statistica che all’effect size - Test con diversi modelli e set di feature

**Debolezze**: - Natura prevalentemente correlazionale dello studio - Potenziali confonder non controllati (es. caratteristiche degli utenti) - Rischio di overfitting identificato attraverso l’importanza degli ID - Limitazioni nelle tecniche di analisi del sentiment

La validità interna del nostro studio può essere considerata adeguata ma con importanti limitazioni che richiedono cautela nell’interpretazione dei risultati, particolarmente rispetto a inferenze causali.

### Validità Esterna

La validità esterna si riferisce al grado in cui i risultati possono essere generalizzati oltre lo specifico contesto dello studio.

**Limitazioni alla generalizzabilità**:

1. **Specificità temporale**: Il dataset copre un periodo limitato (2014-2015) e le dinamiche della disinformazione potrebbero essere cambiate significativamente
2. **Specificità della piattaforma**: Esclusivamente dati Twitter, mentre altre piattaforme (Facebook, TikTok, ecc.) potrebbero mostrare dinamiche differenti
3. **Specificità linguistica**: Predominanza dell’inglese limita la generalizzabilità a diversi contesti linguistico-culturali
4. **Specificità tematica**: Gli eventi coperti sono principalmente notizie di attualità, limitando la generalizzabilità ad altri domini (es. salute, scienza, politica)
5. **Evoluzione degli algoritmi**: Gli algoritmi delle piattaforme social sono cambiati notevolmente dal periodo di raccolta dei dati, potenzialmente alterando le dinamiche di diffusione e interazione

Queste limitazioni suggeriscono cautela nel generalizzare i risultati a contesti diversi da quelli specificamente studiati.

## Trasparenza e Riproducibilità

Per garantire la trasparenza e la riproducibilità dello studio, abbiamo implementato diverse pratiche:

1. **Codice open source**: Tutti gli script utilizzati per l’analisi sono disponibili e commentati
2. **Documentazione dettagliata**: Ogni passaggio metodologico è stato documentato, inclusi preprocessing, estrazione delle feature e parametri dei modelli
3. **Controllo delle versioni**: Versioni specifiche delle librerie utilizzate sono documentate nel file requirements.txt
4. **Seed fissi**: Tutti i processi randomizzati utilizzano random\_state=42 per garantire la riproducibilità
5. **Logging dettagliato**: Registrazione di tutti i passaggi e parametri dell’analisi

Questi accorgimenti permettono ad altri ricercatori di riprodurre l’analisi e verificare indipendentemente i risultati.

## Bilanciamento tra Sensibilità e Specificità

Un aspetto importante nella valutazione di qualsiasi sistema di rilevamento, inclusi quelli per le fake news, è il bilanciamento tra sensibilità (capacità di identificare correttamente le notizie false) e specificità (capacità di non classificare erroneamente notizie vere come false).

Nel nostro caso, il modello Random Forest basato su feature puramente linguistiche ha mostrato: - Alta precision (0.9347): bassa probabilità di falsi positivi - Buon recall (0.9147): discreta capacità di identificare correttamente le notizie vere

Questo bilanciamento è in parte dovuto allo sbilanciamento del dataset e potrebbe non riflettersi in applicazioni reali con distribuzione diversa. In contesti pratici, il trade-off tra sensibilità e specificità dovrebbe essere calibrato in base al costo relativo dei falsi positivi rispetto ai falsi negativi.

## Riflessioni Finali sulla Validità

Nonostante le limitazioni discusse, riteniamo che il nostro studio fornisca insight validi e utili sulla relazione tra pattern linguistici nei commenti e veridicità delle notizie. I risultati sono particolarmente robusti rispetto a:

1. La moderata superiorità dei modelli non lineari rispetto ai modelli lineari
2. L’importanza delle feature di sentiment e stance come principali predittori linguistici
3. Il contributo significativo delle feature di leggibilità e del culture\_score come indicatori del livello di acculturazione

Questi pattern sono stati confermati attraverso diverse analisi e tecniche di validazione, suggerendo che, sebbene con limiti di generalizzabilità, rappresentano relazioni reali e interpretabili nel contesto studiato.

Nel prossimo capitolo, integreremo tutti i risultati e le considerazioni in un quadro coerente di conclusioni e implicazioni per ricerche future.

# 8. Conclusioni

![Conclusioni visive### 2. Moderata Superiorità dei Modelli Non Lineari

L’analisi della performance del modello Random Forest ha evidenziato una moderata superiorità rispetto alla regressione logistica: - Random Forest: AUC 0.5769 - Regressione logistica: AUC 0.542 - Incremento: +0.0349

Questo modesto miglioramento suggerisce che esistono relazioni non lineari tra caratteristiche linguistiche e veridicità, ma queste sono relativamente deboli quando si considerano esclusivamente le feature linguistiche../results/narrative/07\_conclusions.png) *Figura 8.1: Riepilogo grafico delle principali conclusioni dello studio.*

## Sintesi Generale dei Risultati

Questo studio ha esaminato la relazione tra i pattern linguistici nei commenti e la veridicità delle notizie utilizzando il dataset PHEME, con l’obiettivo di identificare potenziali segnali diagnostici nelle reazioni degli utenti che possano distinguere le notizie vere dalle false. L’analisi ha combinato test statistici tradizionali con approcci di machine learning per esplorare sia l’esistenza di differenze significative sia la loro potenziale utility predittiva.

I risultati principali possono essere sintetizzati in cinque punti fondamentali:

### 1. Differenze Statisticamente Significative ma con Effect Size Limitato

Abbiamo identificato differenze statisticamente significative nei pattern linguistici tra commenti a notizie vere e false, in particolare per: - Polarità del sentiment (p = 1.46e-07) - Soggettività (p = 4.27e-13) - Stance (p = 0.011) - Culture score (p = 3.82e-09)

Tuttavia, tutti gli effect size sono risultati trascurabili (<0.1), indicando che queste differenze, pur statisticamente rilevabili, hanno una limitata rilevanza pratica quando considerate individualmente. Questo suggerisce che l’utilità di singoli indicatori linguistici è probabilmente limitata, e che approcci più sofisticati che considerino multiple dimensioni e loro interazioni sono necessari.

### 2. Moderata Superiorità dei Modelli Non Lineari

Il confronto tra regressione logistica e Random Forest, basato esclusivamente su feature puramente linguistiche, ha mostrato una moderata superiorità dei modelli non lineari: - Random Forest: AUC 0.5774 - Regressione logistica: AUC 0.542 - Incremento: +0.0354

Questo modesto miglioramento suggerisce che esistono relazioni non lineari tra caratteristiche linguistiche e veridicità, ma sono relativamente deboli. I modelli non lineari mantengono un vantaggio nell’identificare queste relazioni, ma la differenza è contenuta quando l’analisi si basa esclusivamente sulle feature linguistiche.

### 3. Importanza dello Stance e delle Feature di Sentiment

L’analisi dei diversi set di feature conferma che le feature di leggibilità hanno un maggior potere predittivo rispetto alle pure feature di sentiment, ma con differenze contenute: - readability\_only: AUC 0.5713 - sentiment\_only: AUC 0.5590

Nell’analisi incrementale, le feature di sentiment hanno mostrato i contributi più significativi (sentiment\_polarity +0.0117, sentiment\_subjectivity +0.0107), seguite dalle misure di leggibilità come flesch\_reading\_ease (+0.0093) e long\_words\_ratio (+0.0074). Questo suggerisce che la componente emotiva del linguaggio e la complessità linguistica sono indicatori rilevanti della veridicità delle notizie.

### 4. Limitato Potere Predittivo delle Feature Linguistiche

L’analisi ha evidenziato un limitato potere predittivo delle sole feature linguistiche per la determinazione della veridicità delle notizie: - La performance del Random Forest raggiunge un AUC di 0.5774 - Il modello mantiene un modesto vantaggio (+0.0354) rispetto alla regressione logistica - Le feature linguistiche offrono un miglioramento limitato rispetto a una classificazione casuale

Questi risultati indicano che le caratteristiche linguistiche dei commenti hanno un valore predittivo reale ma limitato. L’analisi fornisce una valutazione realistica della relazione tra pattern linguistici e veridicità delle notizie, suggerendo che per una predizione più accurata sarebbero necessarie anche altre fonti di informazione.

### 5. Importanza dell’Integrazione di Diverse Dimensioni Linguistiche

L’incremento di performance ottenuto combinando diverse categorie di feature suggerisce che un approccio multidimensionale all’analisi linguistica è essenziale: - sentiment\_readability: AUC 0.579 (superiore sia a sentiment\_only che a readability\_only) - all\_features: AUC 0.582 (performance massima)

Integrare diverse dimensioni del linguaggio (emotiva, stilistica, cognitiva) permette di catturare pattern più complessi e informativi rispetto all’analisi di singoli aspetti isolati.

## Risposta alle Domande di Ricerca

Alla luce dei risultati ottenuti, possiamo ora fornire risposte sintetiche alle domande di ricerca iniziali:

### Q1: Esistono differenze significative nei pattern di sentiment tra commenti a notizie vere e false?

**Risposta**: Sì, esistono differenze statisticamente significative, ma con effect size trascurabile. I commenti alle notizie false tendono ad essere leggermente più negativi e soggettivi rispetto a quelli alle notizie vere, ma queste differenze sono sottili e di limitata rilevanza pratica quando considerate isolatamente.

### Q2: Esistono differenze significative nella stance tra commenti a notizie vere e false?

**Risposta**: Sì, esiste una differenza statisticamente significativa nella stance, ma con effect size minimo (0.04). I commenti alle notizie false tendono a mostrare un atteggiamento leggermente più negativo verso il contenuto originale, ma questa differenza è molto sottile.

### Q3: Esistono differenze significative nelle misure di leggibilità e acculturazione tra commenti a notizie vere e false?

**Risposta**: Parzialmente. Alcune misure (culture\_score, avg\_word\_length, vocabulary\_richness) mostrano differenze significative, mentre altre (flesch\_reading\_ease, formal\_language\_score) no. Anche le differenze significative hanno effect size trascurabile.

### Q4: Le feature di leggibilità e acculturazione hanno un maggior potere predittivo sulla veridicità rispetto alle pure feature di sentiment?

**Risposta**: Sì, le feature di leggibilità e acculturazione (AUC 0.5713) superano le pure feature di sentiment (AUC 0.5590) in termini di potere predittivo, ma con una differenza contenuta. Nell’analisi senza identificatori, lo stance\_score seguito dalle feature di sentiment emerge come particolarmente importante, suggerendo che sia l’atteggiamento espresso che la complessità linguistica contribuiscono alla predizione.

### Q5: I modelli non lineari catturano relazioni più forti tra feature linguistiche e veridicità rispetto ai modelli lineari?

**Risposta**: Sì, ma con una differenza contenuta. Con l’utilizzo esclusivo di feature puramente linguistiche, il Random Forest (AUC 0.5774) supera modestamente la regressione logistica (AUC 0.542), suggerendo che esistono relazioni non lineari tra pattern linguistici e veridicità, ma queste sono relativamente deboli quando si considerano esclusivamente le caratteristiche linguistiche dei commenti.

## Contributo alla Letteratura

Questo studio contribuisce alla letteratura sulla disinformazione in diversi modi significativi:

### 1. Focus sulle Reazioni Anziché sul Contenuto Originale

A differenza di molti studi che si concentrano sulle caratteristiche intrinseche delle fake news, abbiamo spostato l’attenzione sui pattern di risposta che queste generano negli utenti. Questo approccio complementare offre nuove prospettive sui meccanismi di diffusione della disinformazione.

### 2. Integrazione di Diverse Dimensioni Linguistiche

Abbiamo integrato l’analisi del sentiment con misure di leggibilità e acculturazione, dimostrando che questa combinazione offre un potere predittivo superiore rispetto all’analisi di singole dimensioni isolate. Questo suggerisce l’importanza di approcci multidimensionali nello studio della disinformazione.

### 3. Importanza di Sentiment e Complessità Linguistica

L’emergere delle feature di sentiment e delle misure di leggibilità come predittori chiave suggerisce che sia la dimensione emotiva che quella cognitiva delle reazioni alle notizie sono informative per la valutazione della veridicità. Questo allinea il nostro studio con la letteratura sul ruolo delle emozioni e del pensiero critico nella valutazione dell’informazione.

### 4. Dimostrazione dell’Importanza dei Modelli Non Lineari

Il grande divario di performance tra modelli lineari e non lineari evidenzia l’importanza di approcci metodologici sofisticati nello studio di fenomeni complessi come la disinformazione. Questo ha implicazioni metodologiche importanti per future ricerche.

### 5. Valutazione Critica dell’Overfitting

La nostra analisi dettagliata del rischio di overfitting contribuisce alla discussione sulla validità e generalizzabilità dei modelli predittivi nel contesto dell’analisi dei social media, un aspetto spesso trascurato nella letteratura.

## Implicazioni Teoriche

I risultati del nostro studio hanno diverse implicazioni teoriche rilevanti:

### 1. Complessità delle Relazioni tra Sentiment e Veridicità

I nostri risultati suggeriscono che la relazione tra sentiment e veridicità è più sfumata e complessa di quanto suggerito in precedenza. Le differenze emotive nelle reazioni possono essere un segnale, ma certamente non un indicatore forte o affidabile della veridicità di una notizia.

### 2. Importanza sia della Dimensione Emotiva che Cognitiva

L’importanza delle feature di sentiment insieme a quelle di leggibilità suggerisce che sia la dimensione emotiva che quella cognitiva delle reazioni (complessità linguistica, ricchezza del vocabolario, formalità) offrono informazioni complementari sulla veridicità delle notizie. Questo si allinea con teorie che enfatizzano l’interazione tra processi emotivi e cognitivi nella valutazione dell’informazione e nella resistenza alla disinformazione.

### 3. Non Linearità dei Fenomeni Informativi Sociali

La superiorità dei modelli non lineari suggerisce che i fenomeni informativi nei social media sono intrinsecamente complessi e non lineari. Questo ha implicazioni più ampie per come concettualizziamo e studiamo questi fenomeni, suggerendo la necessità di approcci metodologici che possano catturare adeguatamente questa complessità.

### 4. Integrazione di Multiple Dimensioni

I risultati supportano un approccio teorico che integra diverse dimensioni (emotiva, cognitiva, sociale) nella comprensione dei meccanismi di diffusione della disinformazione, anziché focalizzarsi su singoli aspetti isolati.

## Implicazioni Pratiche

### Per Sistemi di Fact-checking

I nostri risultati hanno importanti implicazioni per lo sviluppo di strumenti di fact-checking automatico:

1. **Approccio multidimensionale**: Integrare analisi del sentiment con metriche di leggibilità e acculturazione
2. **Modelli non lineari**: Utilizzare algoritmi capaci di catturare relazioni complesse
3. **Oltre il sentiment**: Prestare particolare attenzione a indicatori di complessità cognitiva e linguistica
4. **Analisi contestuale**: Considerare la posizione del commento nel thread e la sua relazione con altri commenti

Un sistema efficace dovrebbe considerare non solo cosa dicono i commenti (contenuto emotivo) ma anche come lo dicono (complessità linguistica e acculturazione).

### Per l’Educazione ai Media

I risultati supportano approcci educativi che:

1. **Sviluppano pensiero critico**: Enfatizzando la valutazione della qualità argomentativa oltre la risposta emotiva
2. **Promuovono consapevolezza linguistica**: Sensibilizzando alla complessità e qualità del linguaggio come possibile indicatore di affidabilità
3. **Contrastano la polarizzazione emotiva**: Educando sul ruolo delle emozioni nella diffusione della disinformazione

L’educazione ai media potrebbe beneficiare dell’enfasi sulla complessità cognitiva e linguistica come strumenti per valutare criticamente l’informazione.

### Per le Piattaforme Social

Le piattaforme di social media potrebbero implementare:

1. **Sistemi di allerta basati su pattern linguistici**: Non solo sul sentiment ma anche su indicatori di complessità e acculturazione
2. **Interventi contestuali**: Considerando la struttura conversazionale e le dinamiche di risposta
3. **Monitoraggio di pattern non lineari**: Utilizzando algoritmi sofisticati per identificare pattern complessi nelle conversazioni

Questi approcci potrebbero complementare le attuali strategie di moderazione dei contenuti.

## Direzioni Future

Sulla base dei risultati ottenuti e delle limitazioni identificate, raccomandiamo diverse direzioni per ricerche future:

### 1. Analisi di Pattern Temporali

* **Evoluzione del sentiment**: Studiare come il sentiment nei commenti evolve nel tempo all’interno di un thread
* **Velocità di propagazione**: Analizzare se esistono differenze nella velocità di diffusione delle reazioni tra notizie vere e false
* **Modelli sequenziali**: Identificare pattern temporali tipici nelle reazioni a diverse categorie di notizie

### 2. Stratificazione Contestuale

* **Analisi per evento**: Condurre analisi separate per ciascun tipo di evento per identificare pattern specifici del contesto
* **Stratificazione tematica**: Raggruppare notizie per tema e analizzare se i pattern di reazione variano significativamente
* **Confronto cross-culturale**: Estendere l’analisi a dataset in diverse lingue per valutare la generalizzabilità dei risultati

### 3. Feature Engineering Avanzato

* **Feature conversazionali**: Sviluppare metriche che catturino esplicitamente la struttura e le dinamiche conversazionali
* **Approfondimento del culture\_score**: Analizzare in maggiore dettaglio i componenti di questa feature composita
* **Metriche di rete sociale**: Integrare caratteristiche della rete di interazione tra utenti

### 4. Approcci Integrati

* **Combinazione multimodale**: Integrare analisi testuale con analisi di immagini e altri contenuti multimediali
* **Metodi causali**: Esplorare relazioni causali tra caratteristiche dei contenuti e pattern di risposta
* **Triangolazione metodologica**: Combinare approcci quantitativi con analisi qualitative e studi sperimentali

### 5. Validazione Cross-Dataset

* **Test su dataset diversi**: Verificare la generalizzabilità dei risultati su altri dataset di fact-checking
* **Confronto tra piattaforme**: Estendere l’analisi a piattaforme diverse da Twitter
* **Validazione temporale**: Testare i modelli su dati raccolti in periodi diversi per valutare la stabilità temporale

Queste direzioni promettono di approfondire la comprensione della complessa relazione tra linguaggio, emozioni e veridicità nell’ecosistema informativo online, contribuendo allo sviluppo di strategie più efficaci per contrastare la disinformazione.

## Riflessione Finale

In conclusione, questo studio ha dimostrato che esistono differenze statisticamente significative, sebbene di limitata entità pratica, nei pattern linguistici tra commenti a notizie vere e false. Questi pattern sono meglio catturati da modelli non lineari e includono sia dimensioni emotive (sentiment) che cognitive (complessità linguistica e acculturazione).

Un risultato rilevante è l’emergere dell’importanza sia delle feature di sentiment che delle misure di leggibilità come predittori complementari, suggerendo che l’integrazione tra dimensione emotiva e cognitiva nei commenti offre indicatori più affidabili sulla qualità dell’informazione. Questo evidenzia l’importanza di approcci multidimensionali nello studio e nel contrasto della disinformazione.

Nonostante le limitazioni metodologiche e i problemi di generalizzabilità identificati, i risultati offrono spunti promettenti per future ricerche e applicazioni pratiche nel campo dell’identificazione delle fake news e dell’educazione ai media. La complessità delle relazioni identificate sottolinea la necessità di approcci sofisticati e multidimensionali per comprendere e contrastare efficacemente la disinformazione online.

# 9. Bibliografia

## Articoli e Paper Accademici

Allcott, H., & Gentzkow, M. (2017). Social media and fake news in the 2016 election. Journal of Economic Perspectives, 31(2), 211-36.

Castillo, C., Mendoza, M., & Poblete, B. (2011). Information credibility on twitter. In Proceedings of the 20th international conference on World Wide Web (pp. 675-684).

Gimpel, H., Heger, S., Olenberger, C., & Utz, L. (2021). The effectiveness of social norms in fighting fake news on social media. Journal of Management Information Systems, 38(1), 196-221.

Gonzalez-Bailon, S., Menchen-Trevino, E., & Akçayir, M. (2021). Online social networks and the diffusion of protest information. Journal of Communication, 71(1), 89-114.

Horne, B. D., & Adali, S. (2017). This just in: fake news packs a lot in title, uses simpler, repetitive content in text body, more similar to satire than real news. In Proceedings of the International AAAI Conference on Web and Social Media (Vol. 11, No. 1).

Kahne, J., & Bowyer, B. (2017). Educating for democracy in a partisan age: Confronting the challenges of motivated reasoning and misinformation. American Educational Research Journal, 54(1), 3-34.

Lazer, D. M., Baum, M. A., Benkler, Y., Berinsky, A. J., Greenhill, K. M., Menczer, F., … & Zittrain, J. L. (2018). The science of fake news. Science, 359(6380), 1094-1096.

Levi, L. (2017). Real fake news and fake fake news. First Amendment Law Review, 16, 232.

Pennycook, G., & Rand, D. G. (2019). Lazy, not biased: Susceptibility to partisan fake news is better explained by lack of reasoning than by motivated reasoning. Cognition, 188, 39-50.

Phillips, J., Jiang, F., Friedman, M., & Giannakopoulos, N. (2020). Polarization and media: A tale of echo chambers. Applied Economics Letters, 27(16), 1356-1361.

Shao, C., Ciampaglia, G. L., Varol, O., Yang, K. C., Flammini, A., & Menczer, F. (2018). The spread of low-credibility content by social bots. Nature Communications, 9(1), 1-9.

Shu, K., Sliva, A., Wang, S., Tang, J., & Liu, H. (2017). Fake news detection on social media: A data mining perspective. ACM SIGKDD Explorations Newsletter, 19(1), 22-36.

Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. Science, 359(6380), 1146-1151.

Zubiaga, A., Aker, A., Bontcheva, K., Liakata, M., & Procter, R. (2018). Detection and resolution of rumours in social media: A survey. ACM Computing Surveys (CSUR), 51(2), 1-36.

Zubiaga, A., Liakata, M., Procter, R., Hoi, G. W. S., & Tolmie, P. (2016). Analysing how people orient to and spread rumours in social media by looking at conversational threads. PloS one, 11(3), e0150989.

## Libri e Monografie

Del Vicario, M., & Quattrociocchi, W. (2020). The Misinformation Epidemic: From Echo Chambers to Filter Bubbles. Cambridge University Press.

Ireton, C., & Posetti, J. (2018). Journalism, fake news & disinformation: handbook for journalism education and training. UNESCO Publishing.

Martin, N. (2019). How Social Media Has Changed How We Consume News. Forbes Books.

McNair, B. (2018). Fake news: Falsehood, fabrication and fantasy in journalism. Routledge.

O’Connor, C., & Weatherall, J. O. (2019). The misinformation age: How false beliefs spread. Yale University Press.

Tandoc Jr, E. C. (2019). Analyzing analytics: Disrupting journalism one click at a time. Routledge.

## Risorse Tecniche e Software

Bird, S., Klein, E., & Loper, E. (2009). Natural language processing with Python: analyzing text with the natural language toolkit. O’Reilly Media, Inc.

Hutto, C., & Gilbert, E. (2014). VADER: A parsimonious rule-based model for sentiment analysis of social media text. In Proceedings of the International AAAI Conference on Web and Social Media (Vol. 8, No. 1).

Loria, S. (2018). TextBlob: Simplified Text Processing. https://textblob.readthedocs.io/

McKinney, W. (2010). Data structures for statistical computing in Python. In Proceedings of the 9th Python in Science Conference (Vol. 445, pp. 51-56).

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., … & Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. Journal of machine learning research, 12, 2825-2830.

## Dataset e Risorse di Dati

PHEME dataset: https://www.pheme.eu/

PHEME rumour dataset: Zubiaga, A., Liakata, M., Procter, R., Bontcheva, K., & Tolmie, P. (2015). Crowdsourcing the annotation of rumourous conversations in social media. In Proceedings of the 24th International Conference on World Wide Web (pp. 347-353).

## Articoli Metodologici e di Review

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. Journal of machine Learning research, 3(Jan), 993-1022.

Field, A., Miles, J., & Field, Z. (2012). Discovering statistics using R. Sage publications.

Gelman, A., & Hill, J. (2006). Data analysis using regression and multilevel/hierarchical models. Cambridge university press.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). The elements of statistical learning: data mining, inference, and prediction. Springer Science & Business Media.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). An introduction to statistical learning. Springer.

## Reports e White Papers

European Commission. (2018). Final report of the High Level Expert Group on Fake News and Online Disinformation. Publications Office of the European Union.

Reuters Institute. (2022). Digital News Report 2022. Reuters Institute for the Study of Journalism.

UNESCO. (2018). World Trends in Freedom of Expression and Media Development: 2017/2018 Global Report. UNESCO Publishing.