



UNIVERSITÀ DEGLI STUDI DI SALERNO

Dipartimento di Informatica

Corso di Laurea Magistrale in Informatica

PROGETTO DI RETI GEOGRAFICHE

Steam Review Insights: Sentiment, Emotion, and Topic Modeling for Game Feedback

GRUPPO
Tommaso Sorrentino

Matricola: 0522501932

Mattia Giella

Matricola: 0522501887

Abstract

Il progetto mira ad analizzare 2 giochi dello stesso genere, Battlefile 2042 e Rainbow Six Siege, utilizzando l'API di Steam per raccogliere recensioni e dati relativi al gradimento dei giochi. L'analisi si concentra su tre tecniche principali: Sentiment Analysis, per determinare il tono complessivo delle recensioni e la percezione dei giocatori; Emotion Analysis, per identificare le emozioni espresse nelle recensioni e comprendere meglio le esperienze emotive degli utenti; e Topic Modeling, per estrarre i temi ricorrenti nelle recensioni, come la giocabilità, la grafica e le meccaniche di gioco. I risultati forniti da queste analisi offriranno una visione dettagliata del feedback degli utenti, supportando decisioni strategiche su come procedere con lo sviluppo del gioco.

Indice

Elenco delle Figure	iii
Elenco delle Tabelle	v
1 Introduzione	1
1.1 Motivazioni	1
1.2 Obiettivi	2
2 Metodologia	4
2.1 Origine dei Dati	4
2.2 Accesso ai Dati delle Recensioni su Steam	5
2.2.1 API Steamworks	5
2.3 Progettazione dello scraper	7
2.3.1 Flusso del Programma	7
2.3.2 Parametri di Default dell'API di Steam	8
2.4 Data Preparation	9
2.4.1 Preparazione dei Dati per Sentiment ed Emotion Analysis . .	10
2.4.2 Preparazione dei Dati per Topic Modelling	11
2.5 Modelli scelti	12
2.5.1 Summarization	12
2.5.2 Sentiment Analysis	13

2.5.3	Emotion Analysis	13
2.5.4	Topic Modeling	13
2.6	Visualizzazione dei dati	15
3	Risultati	16
3.1	Dati Raccolti	16
3.2	Performance dei Modelli di Sentiment ed Emotion Analysis	16
3.3	Paragone fra i giochi	18
3.3.1	Analisi della Distribuzione del Sentiment	18
3.3.2	Confronto	18
3.3.3	Emotion Analysis	19
3.4	Opinioni nel tempo	22
3.4.1	Battlefield 2042	22
3.4.2	Rainbow Six Siege	22
3.5	Risultati della Topic Modeling	23
3.5.1	Analisi del topic modelling	23
4	Conclusioni	28
Bibliografia		30

Elenco delle figure

2.1	Giocatori mensili concorrenti nel periodo [2021-2023][1]	4
2.2	Flowchart dello scraper	8
3.1	Battlefield 2042	17
3.2	Rainbow Six Siege	17
3.3	Iistogrammi che rappresentano l'accuracy del modello di sentiment analysis	17
3.4	Battlefield 2042	18
3.5	Rainbow Six Siege	18
3.6	Iistogrammi che rappresentano l'accuracy del modello di emotion analysis	18
3.7	Battlefield 2042	19
3.8	Rainbow Six Siege	19
3.9	Iistogrammi che rappresentano la distribuzione del sentiment	19
3.10	Battlefield 2042	19
3.11	Rainbow Six Siege	19
3.12	Piechart della distribuzione della sentiment analysis	19
3.13	Battlefield 2042	21
3.14	Rainbow Six Siege	21
3.15	Iistogrammi che rappresentano la distribuzione dell'emotion	21

3.16 Battlefield 2042	21
3.17 Rainbow Six Siege	21
3.18 Piechart dell'emotion distribution	21
3.19 Battlefield 2042	22
3.20 Rainbow Six Siege	22
3.21 Heatmap delle emozioni per sentiment	22
3.22 Variazione dello score del sentiment nel tempo per Battlefield 2042 .	23
3.23 Variazione dello score del sentiment nel tempo per Rainbow Six Siege	23
3.24 Primo Topic Battlefield 2042	24
3.25 Primo Topic Rainbow Six Siege	25
3.26 Battlefield 2042	25
3.27 Secondo Topic Battlefield 2042	25
3.28 Terzo Topic Battlefield 2042	26
3.29 Quarto Topic Rainbow Six Siege	27

Elenco delle tabelle

CAPITOLO 1

Introduzione

Il mercato videoludico è oggi uno dei settori più dinamici e in rapida espansione dell'industria dell'intrattenimento. Nel 2023, il mercato globale dei videogiochi ha generato ricavi per circa 184 miliardi di dollari, segnando una crescita dello 0,6% rispetto all'anno precedente [2]. Questa cifra supera i ricavi combinati di altri settori dell'intrattenimento, come lo streaming video, la musica e il box office cinematografico [3].

In un contesto così competitivo, il successo di un videogioco dipende da una combinazione di fattori, tra cui la qualità tecnica, il gameplay, la gestione delle aspettative dei giocatori e l'efficacia della comunicazione con il pubblico. Mentre alcuni titoli riescono a conquistare il mercato e a costruire una solida community di fan, altri falliscono nel rispondere alle aspettative, subendo un forte contraccolpo in termini di vendite e reputazione.

1.1 Motivazioni

Il nostro progetto nasce dall'esigenza di capire meglio le dinamiche che determinano il successo o il fallimento di un videogioco, partendo dalle opinioni degli utenti. Nel mercato videoludico, dove le recensioni e i feedback online giocano un

ruolo cruciale, è importante analizzare come i giocatori percepiscono un titolo e quali aspetti influenzano le loro valutazioni.

Abbiamo scelto di concentrarci su *Battlefield 2042* e *Rainbow Six Siege* perché rappresentano due casi opposti: il primo è stato accolto in modo critico dal pubblico, mentre il secondo ha ottenuto un consenso crescente nel tempo. Questi due esempi ci permettono di esplorare i fattori chiave che guidano il successo o l'insuccesso di un gioco.

Analizzare le recensioni degli utenti ci offre una base concreta per identificare le opinioni comuni, i problemi più sentiti e le caratteristiche apprezzate. Questo tipo di analisi può essere utile non solo per confrontare i due titoli, ma anche per fornire spunti pratici per le aziende interessate a migliorare i propri prodotti.

1.2 Obiettivi

Questo progetto si propone di analizzare due titoli: *Battlefield 2042* e *Rainbow Six Siege*. Utilizzando tecniche di **topic modelling**, **sentiment analysis** e **emotion analysis**, intendiamo esplorare le recensioni degli utenti, al fine di comprendere le ragioni del diverso impatto sul mercato di questi giochi. In particolare, ci proponiamo di:

- **Analizzare il sentimento delle recensioni** per quantificare il livello di soddisfazione o insoddisfazione (espresse con uno score compreso tra 1-5) dagli utenti.
- **Effettuare un'analisi delle emozioni** per individuare le emozioni prevalenti (scelte fra rabbia, disgusto, paura, gioia, tristezza, sorpreso o neutralità) espresse dagli utenti nei loro feedback.
- **Identificare i temi principali** (topics) emersi dalle recensioni degli utenti, per scoprire quali aspetti dei giochi sono stati maggiormente apprezzati o criticati.
- **Mettere a confronto i risultati tra i due titoli**, evidenziando i fattori chiave che hanno influenzato il successo di *Rainbow Six Siege* rispetto alle criticità riscontrate in *Battlefield 2042*.

L’obiettivo è duplice: da un lato, identificare i temi e le emozioni principali che caratterizzano il feedback degli utenti; dall’altro, fornire un’analisi comparativa che possa essere utile sia per i produttori di videogiochi sia per gli analisti di mercato.

Per condurre l’analisi, utilizzeremo dati provenienti da *Steam*, una delle principali piattaforme di distribuzione digitale e gestione dei videogiochi a livello globale. *Steam*, sviluppata da Valve Corporation, è considerata il "**best-in-class**" nel settore grazie alla sua vasta base di utenti, che supera i 120 milioni di giocatori attivi mensili [4], e alla ricchezza delle recensioni pubblicate dagli utenti, che rappresentano una risorsa preziosa per comprendere le opinioni e le emozioni del pubblico.

CAPITOLO 2

Metodologia

2.1 Origine dei Dati

I dati utilizzati per il nostro progetto provengono dalla piattaforma *Steam*, uno dei principali marketplace e social network per videogiocatori. Steam è stata scelta per la sua ampia base di utenti attivi e per la disponibilità di recensioni dettagliate, che rappresentano un'ottima risorsa per analizzare opinioni e sentimenti relativi ai giochi.

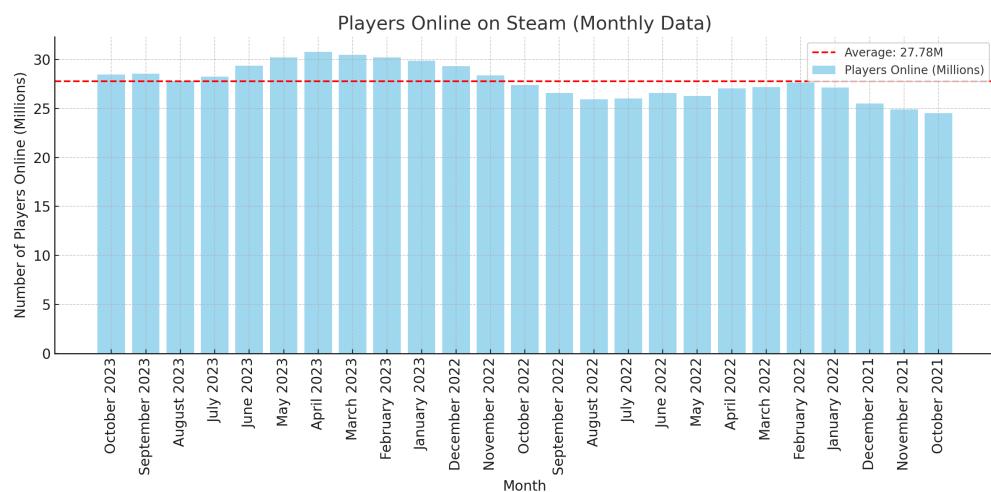


Figura 2.1: Giocatori mensili concorrenti nel periodo [2021-2023][1]

2.2 Accesso ai Dati delle Recensioni su Steam

Per raccogliere i dati delle recensioni degli utenti dalla piattaforma *Steam*, utilizziamo l'API delle recensioni di Steamworks. Questo strumento consente di ottenere un dump strutturato in formato JSON delle recensioni, offrendo una vasta gamma di parametri configurabili per personalizzare la richiesta in base alle esigenze dell'analisi.

2.2.1 API Steamworks

Il metodo base per accedere alle recensioni è il seguente:

```
GET store.steampowered.com/appreviews/<appid>?json=1
```

Dove `<appid>` rappresenta l'identificativo unico dell'applicazione di cui si vogliono recuperare le recensioni.

Principali Parametri

- **filter:** Consente di specificare il criterio di ordinamento delle recensioni:
 - `recent`: recensioni ordinate per data di creazione.
 - `updated`: recensioni ordinate per ultimo aggiornamento.
 - `all`: predefinito, ordinamento per utilità.
- **language:** Specifica la lingua delle recensioni da recuperare. Può essere impostato su `all` per ottenere recensioni in tutte le lingue.
- **day_range:** Definisce un intervallo temporale, da ora fino a un massimo di 365 giorni prima, per filtrare le recensioni.
- **cursor:** Utilizzato per paginare le richieste. Ogni risposta include un valore di `cursor` per richiedere il gruppo successivo.
- **review_type e purchase_type:** Filtrano rispettivamente le recensioni positive/negative e quelle scritte da utenti che hanno acquistato il gioco su Steam o altrove.

- **num_per_page**: Specifica il numero di recensioni restituite per richiesta (massimo 100).
- **filter_offtopic_activity**: Le recensioni fuori tema (anche note come "review bomb") sono scartate e non vengono restituite in questa API.

Output dei Dati

L'API restituisce un elenco JSON contenente:

- **reviews**: Lista delle recensioni con dettagli come testo, lingua, data di creazione e metadati relativi all'utente (es. ore giocate).
- **query_summary**: Informazioni aggregate, come il numero totale di recensioni positive e negative.

Esempio di una parte del JSON restituito:

```
{
  "success": 1,
  "query_summary": {
    "num_reviews": 20,
    "total_positive": 1500,
    "total_negative": 500
  },
  "reviews": [
    {
      "recommendationid": "12345",
      "author": {
        "steamid": "123456789",
        "num_games_owned": 50,
        "playtime_forever": 120
      },
      "review": "Amazing game! Highly recommend.",
      "voted_up": true,
    }
  ]
}
```

```
    "timestamp_created": 1693747200
}
]
}
```

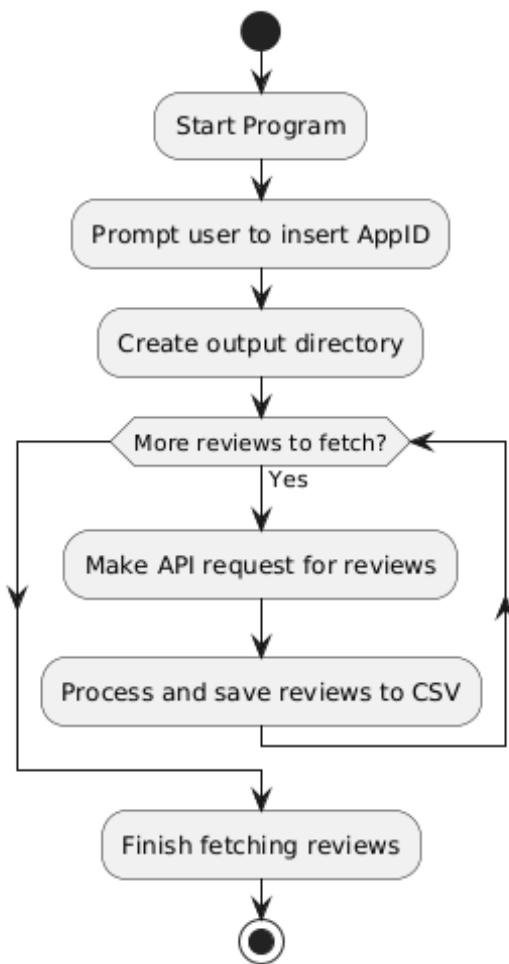
2.3 Progettazione dello scraper

Lo scraper è progettato per raccogliere recensioni di giochi dalla piattaforma *Steam* utilizzando l'API fornita da Steamworks. Il funzionamento dello scraper può essere suddiviso in diverse fasi, come descritto di seguito.

2.3.1 Flusso del Programma

Il flusso del programma segue una sequenza chiara, illustrata anche nel diagramma di flusso della figura 2.2.

1. **Inserimento dell'AppId:** L'utente può inserire un AppID specifico, corrispondente al gioco di cui desidera raccogliere le recensioni. Il programma verifica l'AppID fornito utilizzando l'API di Steam per assicurarsi che i dettagli del gioco siano corretti.
2. **Creazione della directory di output:** Dopo aver selezionato la modalità e confermato l'AppID, viene creata una directory per salvare le recensioni in formato CSV.
3. **Raccolta delle recensioni:** Il programma iterativamente effettua una richiesta all'API per ottenere un gruppo (*batch*) di recensioni, salvate in un file CSV.
4. **Conclusione:** Al termine della raccolta, il programma si interrompe, segnalando che tutte le recensioni sono state salvate correttamente.

**Figura 2.2:** Flowchart dello scraper

2.3.2 Parametri di Default dell'API di Steam

Lo scraper utilizza una serie di parametri di default per configurare le richieste all'API di Steam, la configurazione dei parametri scelta consente di ottenere tutte le recensioni in lingua inglese e di ottimizzare le richieste per grandi volumi di dati.

- **json**: Impostato a 1, questo parametro specifica che la risposta dell'API deve essere restituita in formato JSON, che è facilmente leggibile e processabile dal programma.
- **filter**: Impostato a "recent", questo parametro configura l'API per restituire tutte le recensioni, ordinate in base alla data di creazione.
- **language**: Impostato a "english", limita le recensioni restituite a quelle scritte in lingua inglese. Questo parametro è importante per garantire che le

recensioni siano comprensibili e coerenti con l’analisi prevista.

- **num_per_page:** Impostato a 100, specifica il numero massimo di recensioni da restituire per ogni richiesta. Questo valore, che è il massimo consentito dall’API, ottimizza il numero di chiamate necessarie per ottenere un dataset completo.
- **purchase_type:** Impostato a "all", permette di includere recensioni da tutte le tipologie di acquisto, sia effettuate direttamente su Steam ("steam") sia tramite altri metodi ("non_steam_purchase"). Questo parametro garantisce una visione più completa delle opinioni degli utenti.
- **filter_offtopic_activity:** Per impostazione predefinita, questo parametro scarta le recensioni fuori tema, comunemente note come *review bomb*, che possono influenzare negativamente il punteggio generale senza essere rilevanti per la qualità effettiva del gioco.

2.4 Data Preparation

I dati raccolti sono stati sottoposti a un processo di **data engineering** per garantirne la qualità e prepararli adeguatamente all’analisi. Avendo raccolto i dati con la SteamReview API, le recensioni sono state filtrate in lingua inglese e sono state automaticamente rimosse le recensioni non significative.

- **Rimozione delle recensioni non significative:** Sono state eliminate recensioni associate a fenomeni di *review bombing*. Queste recensioni, spesso non costruttive e scritte per scopi non legati alla qualità del gioco, includono commenti fuori tema come:

"Don't buy this game! The company supports unfair policies!"

oppure recensioni estremamente negative e non correlate all’esperienza di gioco, come:

"This game is really bad because of the developers' political stance!"

La rimozione di tali recensioni è stata effettuata per ridurre il rumore nei dati e garantire un'analisi basata su opinioni autentiche e pertinenti.

Di seguito, il processo di preparazione è stato adattato in base ai requisiti specifici delle analisi di **Sentiment ed Emotion** e **Topic Modelling**.

2.4.1 Preparazione dei Dati per Sentiment ed Emotion Analysis

Il processo di preparazione per le analisi di sentimento ed emozioni si è concentrato sull'ottimizzazione dei dati senza perdere informazioni rilevanti:

- **Rimozione della punteggiatura:** Non effettuata, poiché potrebbe comportare perdita di informazioni. Ad esempio:

"Amazing game! Loved the graphics and gameplay."

contiene emozioni fortemente legate all'uso della punteggiatura ("!").

- **Lematizzazione:** Non applicata, poiché la tokenizzazione di modelli avanzati come *BERT* gestisce già in modo efficace la normalizzazione. Per esempio, termini come:

"playing" e *"played"*

vengono interpretati correttamente senza necessità di ridurli a *"play"*.

- **Riassunto del testo (Summarization):** Abbiamo utilizzato un modello bart per ridurre la lunghezza dell'input, rendendo l'elaborazione più efficiente. Ad esempio, una recensione come:

"The game is very well made, with stunning visuals and engaging gameplay, but the story feels a bit rushed."

viene sintetizzata in:

"Great visuals and gameplay, rushed story."

```

1 summary = summarizer(review, max_length=150,
2 min_length=20, do_sample=False)

```

Listing 1: Snippet di codice per il riassunto delle recensioni.

2.4.2 Preparazione dei Dati per Topic Modelling

Per il **Topic Modelling**, sono stati adottati metodi più tradizionali di pre-elaborazione del testo:

- **Rimozione della punteggiatura:** Effettuata per semplificare la tokenizzazione.

Ad esempio, il testo:

"This game is fantastic! A masterpiece."

diventa:

"This game is fantastic A masterpiece"

- **Rimozione delle stopwords generiche:** Parole comuni come *"the"*, *"is"*, *"and"* sono state rimosse utilizzando la libreria *nltk*. Ad esempio:

"The game is fantastic and engaging."

viene ridotto a:

"game fantastic engaging"

- **Lemmatizzazione:** Applicata con *nltk* (*wordnet*) per ridurre le parole alla loro forma base. Ad esempio:

"playing" e *"played"*

vengono normalizzati a:

"play"

- **Rimozione di stopwords di dominio:** Termini generici come *"game"* che possono dominare i topic sono stati rimossi. Ad esempio:

"The game offers great gameplay and game mechanics."

viene ridotto a:

"offers great gameplay mechanics"

```

1 cleaned_reviews = [
2     ' '.join([
3         lemmatizer.lemmatize(word)
4         for word in re.sub(r'[^\w\s]', ' ', review.lower()).split()
5         if word not in stop_words
6     ])
7     for review in reviews
8 ]

```

Listing 2: Esempio di codice per rimozione punteggiatura, stopwords e lemmatizzazione.

2.5 Modelli scelti

Per il nostro progetto, abbiamo selezionato una combinazione di modelli e tecniche avanzate per analizzare le recensioni dei giochi, tenendo conto della qualità dei dati e degli obiettivi specifici di ciascuna analisi.

2.5.1 Summarization

Abbiamo utilizzato il modello **facebook/bart-large-cnn**, un modello pre-addestrato basato sull'architettura BART (Bidirectional and Auto-Regressive Transformers), specificamente ottimizzato per il riassunto di testi. Questo modello è particolarmente adatto al nostro progetto perché:

- È in grado di ridurre recensioni lunghe e complesse in versioni più concise, preservando le informazioni chiave.
- Aiuta a gestire recensioni molto dettagliate che potrebbero contenere informazioni ridondanti, focalizzando l'analisi su aspetti essenziali come sentimenti ed emozioni.

- Supporta un input eterogeneo, come il linguaggio naturale utilizzato nelle recensioni, che spesso include un mix di opinioni personali e descrizioni tecniche.

2.5.2 Sentiment Analysis

Per l’analisi del sentiment, abbiamo scelto il modello **nlptown/bert-base-multilingual-uncased-sentiment**. Questo modello, basato sull’architettura BERT (Bidirectional Encoder Representations from Transformers), è addestrato su recensioni multilingue, rendendolo altamente versatile. La sua adeguatezza al nostro progetto si basa su:

- La capacità di classificare il sentiment su una scala da 1 a 5, che permette di quantificare con precisione il livello di soddisfazione espresso dagli utenti.
- La gestione robusta di testi con variazioni linguistiche, comuni nelle recensioni di giochi, anche quando scritte in inglese non standard.

2.5.3 Emotion Analysis

Per identificare le emozioni predominanti nelle recensioni, abbiamo scelto il modello **j-hartmann/emotion-english-distilroberta-base**. Basato sull’architettura DistilRoBERTa, questo modello è stato addestrato su dati annotati per il riconoscimento delle emozioni. È particolarmente adatto al nostro progetto perché:

- Permette di rilevare emozioni come gioia, tristezza, rabbia, disgusto, paura e sorpresa, che sono rilevanti per comprendere le reazioni degli utenti ai giochi.
- È ottimizzato per la lingua inglese, garantendo un’alta precisione su testi naturali come le recensioni.

2.5.4 Topic Modeling

Per il topic modelling, abbiamo scelto una combinazione di tecniche ed efficaci, che si integrano bene con il nostro dataset e i nostri obiettivi.

Term Frequency: Term Frequency-Inverse Document Frequency

Per calcolare la frequenza dei termini, abbiamo utilizzato il metodo **TfidfVectorizer** (Term Frequency-Inverse Document Frequency), una tecnica ampiamente utilizzata nel processamento del linguaggio naturale (NLP). Questa tecnica è composta da due componenti principali:

- **Term Frequency (TF)**: Calcola la frequenza di un termine specifico (t) in un documento (d)¹. Viene definita come:

$$TF(t, d) = \frac{\text{Numero di occorrenze di } t \text{ in } d}{\text{Numero totale di termini in } d}$$

Questo valore misura quanto un termine sia rilevante per un documento specifico.

- **Inverse Document Frequency (IDF)**: Ridimensiona il peso del termine (t) in base alla sua rarità nel corpus (N). Viene definita come:

$$IDF(t) = \log \frac{N}{1 + n_t}$$

Dove n_t è il numero di documenti che contengono il termine t . I termini comuni in molti documenti ricevono un peso minore.

La **matrice TF-IDF** generata contiene i pesi dei termini per ciascun documento ed è utilizzata come input per il modello di **Latent Dirichlet Allocation (LDA)**. Questo approccio consente di identificare i temi latenti nel corpus in modo più preciso, riducendo l'influenza di termini molto comuni o poco rilevanti.

Topic Modelling: Latent Dirichlet Allocation (LDA)

Per identificare i temi ricorrenti nelle recensioni, abbiamo utilizzato **Latent Dirichlet Allocation (LDA)**, una delle tecniche più consolidate per il topic modelling. La scelta di LDA è giustificata da:

- La capacità di scoprire temi latenti all'interno del testo, che possono fornire insight preziosi sugli aspetti più discussi dai giocatori.

¹Un documento è una singola unità testuale all'interno di un corpus. Può essere una recensione, un articolo, una frase, o qualsiasi altro testo coerente trattato come un'entità separata per l'analisi.

- L'efficacia nell'identificare cluster semantici basati sulle parole più rilevanti, semplificando l'interpretazione dei risultati.
- La possibilità di regolare il numero di topic, adattandolo alle esigenze del nostro dataset e alle domande di ricerca.

2.6 Visualizzazione dei dati

I risultati ottenuti sono stati rappresentati attraverso tecniche di **data visualization** per facilitare la comprensione e l'interpretazione dei dati. In particolare, abbiamo utilizzato:

- Grafici a barre e a torta per la distribuzione del sentiment e dell'emoticon.
- Word cloud per rappresentare i termini più frequenti emersi nei topic.
- Heatmap per evidenziare la correlazione tra emozioni e score.

Le visualizzazioni sono state create con librerie Python come *Matplotlib*, *Seaborn* e *Plotly*.

CAPITOLO 3

Risultati

3.1 Dati Raccolti

I dati raccolti dallo scraper ammontano a circa 600.000 recensioni per Rainbow Six Siege e circa 200.000 per Battlefield 2042. I dati sono stati in seguito analizzati dai vari modelli, di seguito analizzeremo i risultati di queste analisi.

3.2 Performance dei Modelli di Sentiment ed Emotion Analysis

I grafici presentano i risultati relativi all'accuracy dei modelli applicati per la sentiment analysis e l'emotion analysis nei giochi *Battlefield 2042* e *Rainbow Six Siege*.

Sentiment Analysis

I modelli faticano maggiormente con le recensioni intermedie (2-3 stelle), come si può osservare dalle immagini 3.3, probabilmente a causa dell'ambiguità intrinseca delle valutazioni medie.

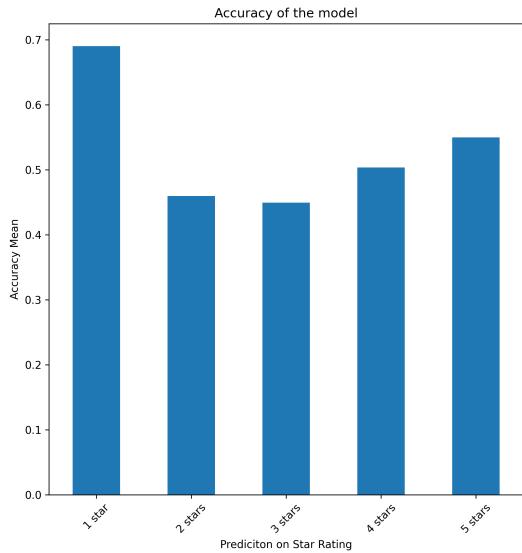


Figura 3.1: Battlefield 2042

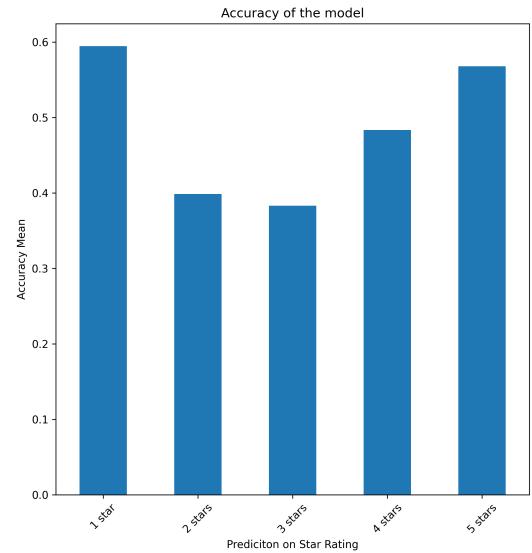
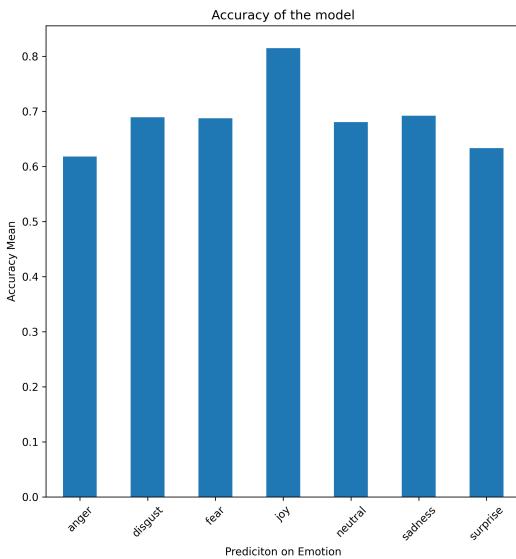
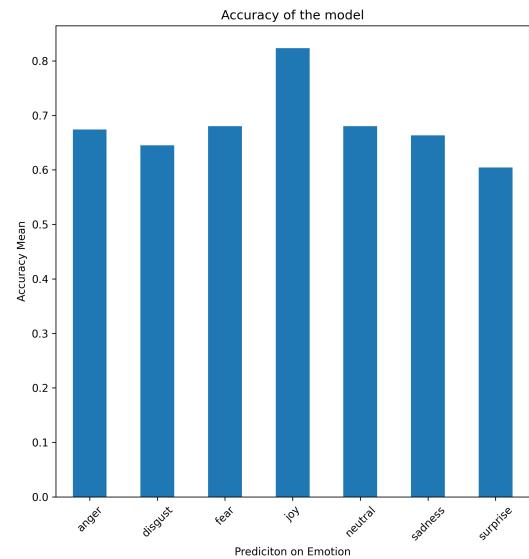


Figura 3.2: Rainbow Six Siege

Figura 3.3: Istogrammi che rappresentano l'accuracy del modello di sentiment analysis

Emotion Analysis

Come è evidenziato nell'immagine 3.6, alcune emozioni potrebbero essere espresse attraverso termini, espressioni o pattern linguistici più facilmente individuabili dal modello. Parole chiave fortemente correlate ad un'emozione (es “joy”, termini di lode o entusiasmo per “gioia”) rendono la classificazione più agevole. Al contrario, emozioni più sfumate come “surprise” potrebbero essere manifestate con termini più ambigui, creando maggiore incertezza nella classificazione.

**Figura 3.4:** Battlefield 2042**Figura 3.5:** Rainbow Six Siege**Figura 3.6:** Istogrammi che rappresentano l'accuracy del modello di emotion analysis

3.3 Paragone fra i giochi

3.3.1 Analisi della Distribuzione del Sentiment

I grafici rappresentano la distribuzione delle valutazioni degli utenti (da 1 a 5 stelle) per *Battlefield 2042* e *Rainbow Six Siege*. Questi risultati evidenziano differenze significative nel sentiment generale percepito.

3.3.2 Confronto

Come si evince dai grafici, *Battlefield 2042* è caratterizzato da un sentiment fortemente negativo e polarizzato, caratterizzato da una predominanza di valutazioni negative, con il 40% delle recensioni a 1 stella (più di 40000 come mostrato nella figura 3.7). Le recensioni positive (4 e 5 stelle) costituiscono solo il 32.8%, riflettendo un feedback prevalentemente insoddisfatto. In contrapposizione, *Rainbow Six Siege* gode di un maggiore consenso positivo, con il 34.7% delle recensioni a 5 stelle (corrispondente a circa 200000 come mostrato in figura 3.8) e il 26.2% a 4 stelle (con quasi 150000 recensioni); le recensioni a 1 stella corrispondono solo il 20.5%.

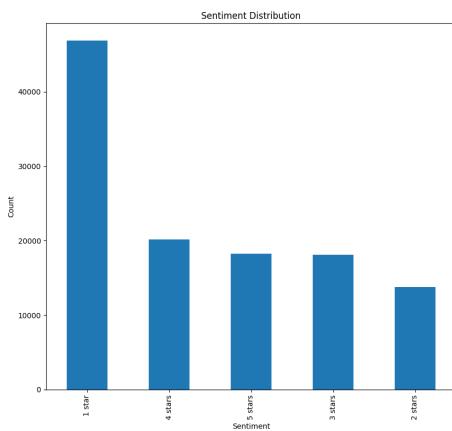
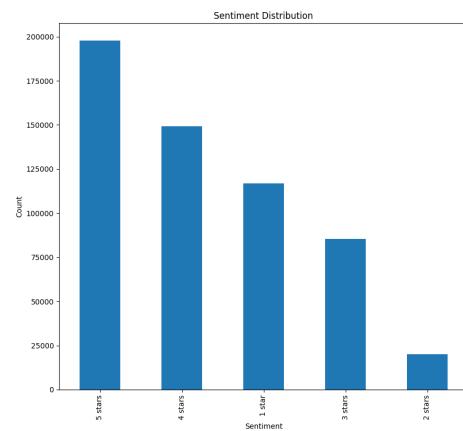

Figura 3.7: Battlefield 2042

Figura 3.8: Rainbow Six Siege

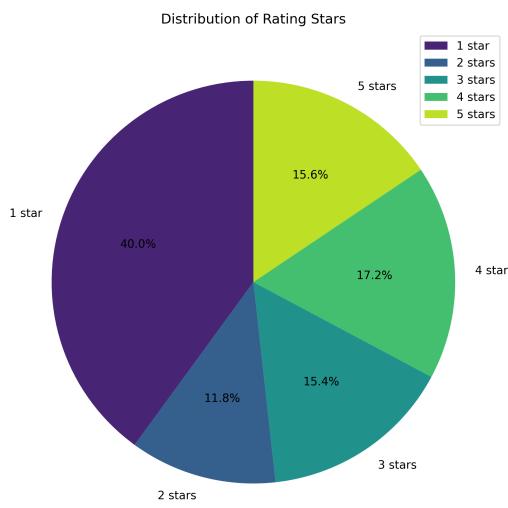
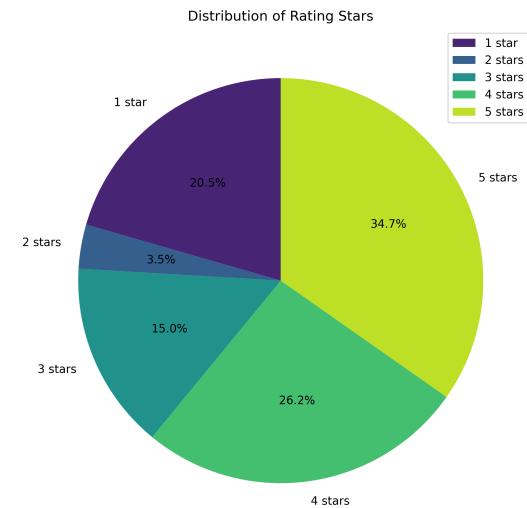
Figura 3.9: Istogrammi che rappresentano la distribuzione del sentiment

Figura 3.10: Battlefield 2042

Figura 3.11: Rainbow Six Siege

Figura 3.12: Piechart della distribuzione della sentiment analysis

3.3.3 Emotion Analysis

Analizzando i risultati dei grafici, emergono alcuni punti chiave che mettono in luce differenze significative tra i due giochi:

- **Neutralità:** Entrambi i giochi presentano una predominanza di recensioni neutre, con una percentuale più alta per *Rainbow Six Siege* (39% contro 32.6%), il che potrebbe indicare una maggiore stabilità emotiva nei feedback degli utenti.

- **Gioia:** *Rainbow Six Siege* registra una percentuale dell'emozione "joy" (31.6%), circa 170.000 recensioni come osservabile in figura 3.15, più alta rispetto a *Battlefield 2042* (20.7%), evidenziando un maggiore gradimento generale.
- **Emozioni negative:** Le recensioni di *Battlefield 2042* sono più caratterizzate da emozioni negative come **disgusto** e **tristezza**, mentre *Rainbow Six Siege* mostra valori più contenuti per queste emozioni.
- **Emozioni marginali:** Entrambi i giochi hanno percentuali molto basse per emozioni come **paura** e **sorpresa**, suggerendo che questi sentimenti non sono prevalenti tra gli utenti.

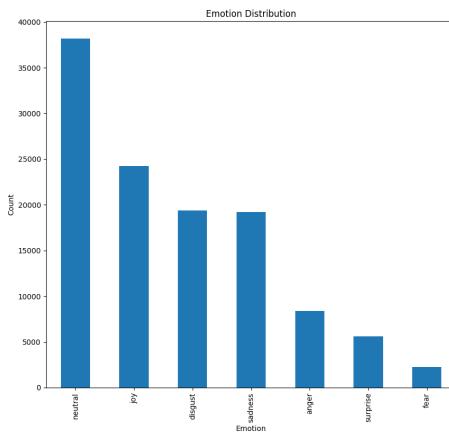


Figura 3.13: Battlefield 2042

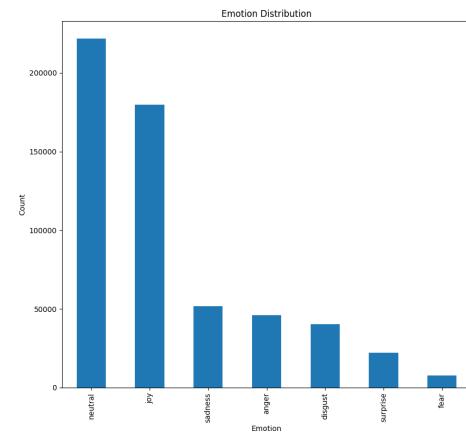


Figura 3.14: Rainbow Six Siege

Figura 3.15: Istogrammi che rappresentano la distribuzione dell'emotion

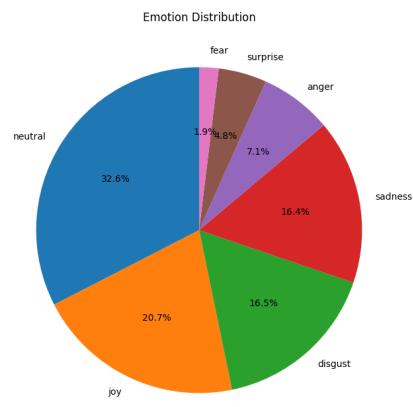


Figura 3.16: Battlefield 2042

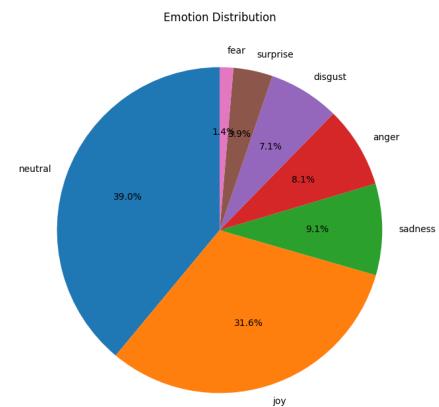


Figura 3.17: Rainbow Six Siege

Figura 3.18: Piechart dell'emotion distribution

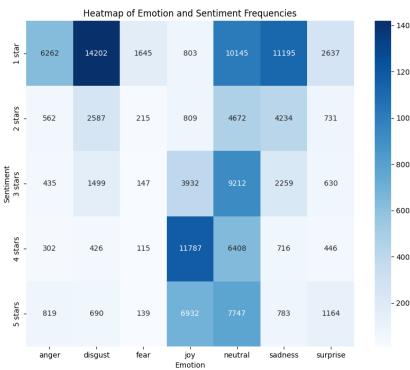
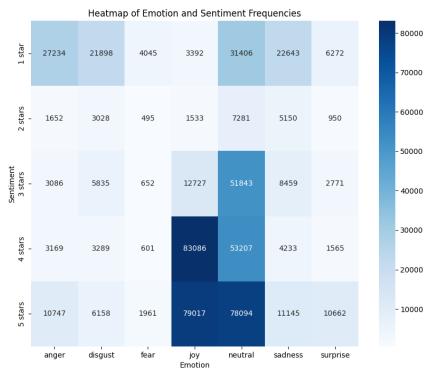

Figura 3.19: Battlefield 2042

Figura 3.20: Rainbow Six Siege

Figura 3.21: Heatmap delle emozioni per sentiment

3.4 Opinioni nel tempo

3.4.1 Battlefield 2042

Dai risultati dell'emotion e del sentiment, si è evidenziata una media dell'andamento decisamente più negativo di Battlefield 2042 rispetto a Rainbow Six Siege. Analizzando il valore nel sentiment possiamo notare che il periodo peggiore per il gioco è stato il lancio, evidenziato da una media di stelle fra 2 e 2,5 per i primi 5/6 mesi. Il gioco ha poi ricevuto recensioni leggermente più positive nel periodo successivo, però rimanendo su un punteggio mediocre compreso fra 3 e 3,5.

3.4.2 Rainbow Six Siege

Il grafico 3.23 mostra un trend complessivamente in diminuzione nel corso degli anni con una forte variabilità nel tempo. Mentre nel periodo tra il 2019 e il 2020 i valori medi oscillano per lo più tra 3,4 e 3,8, con un picco che si attesta fra 3,6 e 3,9 nel periodo di fine 2019, inizio 2020. Col passare degli anni, la media tende progressivamente ad abbassarsi, stabilizzandosi su valori più vicini a 3,2 - 3,4 nel periodo che va dal 2021 al 2024.

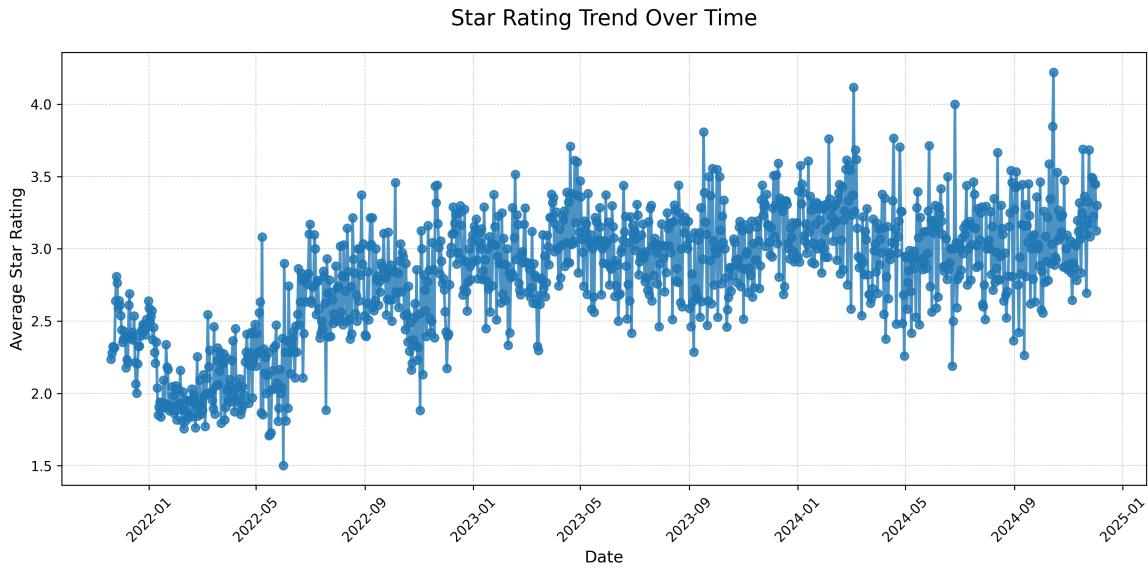


Figura 3.22: Variazione dello score del sentimento nel tempo per Battlefield 2042

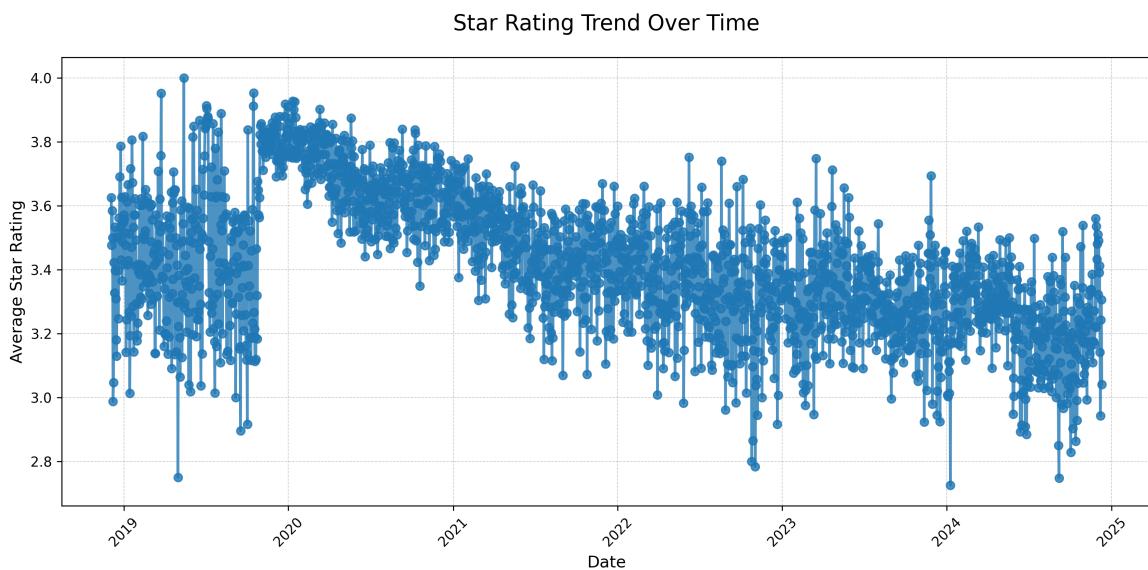


Figura 3.23: Variazione dello score del sentimento nel tempo per Rainbow Six Siege

3.5 Risultati della Topic Modeling

Di seguito siamo andati ad analizzare i risultati dei topic più significativi.

3.5.1 Analisi del topic modelling

Nei grafici 3.24, la parola “bug” è chiaramente tra le più frequenti, così come “issue”. Ciò suggerisce che gli utenti discutono attivamente di problemi tecnici,

evidenziando che stabilità e funzionamento del gioco rappresentano un aspetto centrale delle conversazioni. L'enfasi sui bug indica frustrazione e la necessità di interventi correttivi.

Il termine "fun" risulta molto ricorrente, affiancato da parole come "feel" e "better". Questi termini suggeriscono che, nonostante i problemi tecnici, i giocatori percepiscano o desiderino percepire un'esperienza di gioco divertente e appagante. La presenza di "fun" indica dunque che c'è un potenziale positivo nel gameplay, che però potrebbe essere offuscato dai problemi. Le parole "launch", "release" e "update" rimandano a momenti specifici del ciclo di vita del gioco, come il lancio iniziale e i successivi aggiornamenti. Ciò indica che le conversazioni non si limitano alla sola esperienza di gioco attuale, ma includono commenti sul passato (il lancio del titolo) e sulle aspettative future (gli update). Gli utenti sembrano dunque interessati alla direzione in cui il gioco sta andando e all'impatto degli aggiornamenti sulla qualità complessiva. Se da un lato "bug" e "issue" sottolineano le criticità, parole come "better", "feel" e "fun" suggeriscono un desiderio di vedere il gioco migliorare. Questo scenario indica che il topic non è interamente negativo, bensì caratterizzato da un mix di frustrazione e speranza.

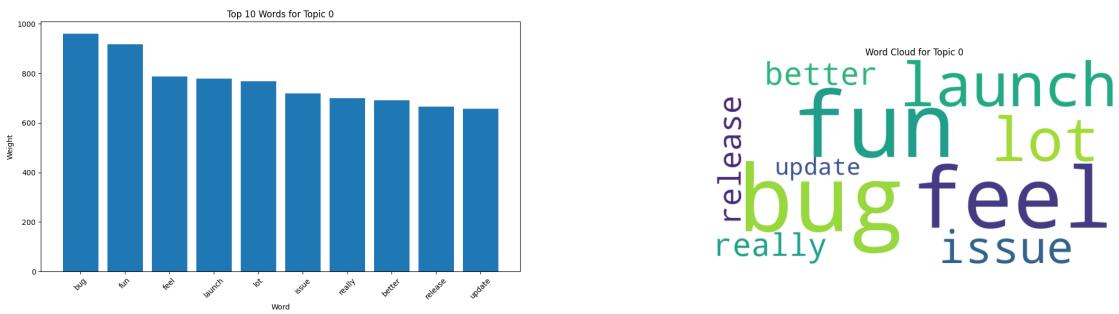


Figura 3.24: Primo Topic Battlefield 2042

Dalla word cloud e dall'istogramma 3.25 emergono temi chiave che evidenziano sia critiche che aree di interesse per i giocatori. Il termine "suck" spicca, indicando una forte insoddisfazione legata a vari aspetti del gioco, probabilmente l'esperienza complessiva, in particolare volta agli "operator", spesso al centro di discussioni riguardanti bilanciamento o nuove aggiunte. "Shoot" e "gameplay" confermano l'attenzione sui meccanismi centrali del gioco, ma il tono generale delle parole lascia

intuire più critiche che apprezzamenti. Infine, termini come "community" e "new" riflettono sia il dibattito su nuovi contenuti sia possibili problematiche legate alla tossicità o alla gestione del gioco.

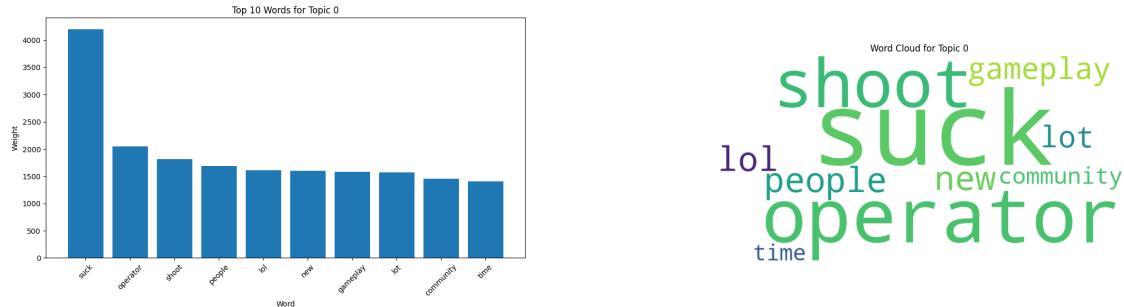


Figura 3.25: Primo Topic Rainbow Six Siege

Il termine "bad" domina il topic 3.27, indicando un'ampia insoddisfazione. Parole come "map", "vehicle", e "server" compaiono tra i termini più frequenti, suggerendo che le critiche siano legate a questi aspetti del gioco. Anche "gun" e "feel" potrebbero indicare problemi percepiti nei controlli delle armi o nell'esperienza immersiva. Inoltre le recensioni sembrano essere polarizzate. Da una parte, ci sono critiche significative (es. "bad", "server"), dall'altra, parole come "love" suggeriscono che alcuni utenti trovino comunque aspetti del gioco piacevoli. Questi dati indicano la necessità di approfondire i temi critici, in particolare relativi a stabilità, ma anche di valorizzare gli elementi positivi emersi.

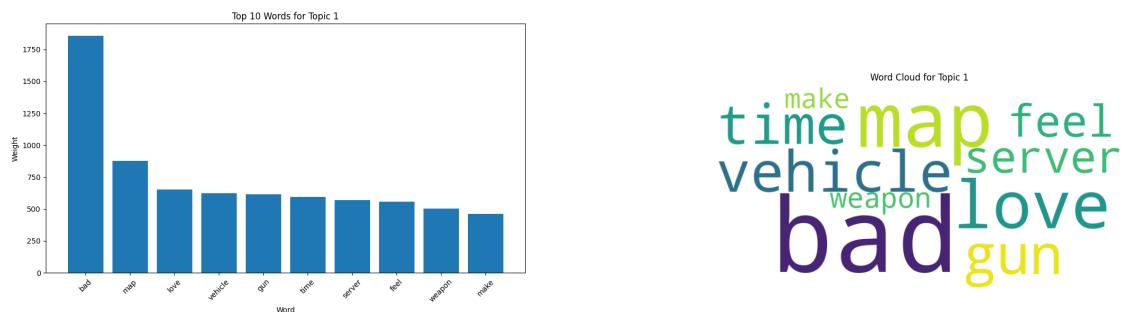


Figura 3.26: Battlefield 2042

Figura 3.27: Secondo Topic Battlefield 2042

Parole altamente negative domininano il topic 3.28. "Trash" (spazzatura) e "money" sono le più frequenti, suggerendo una percezione estremamente negativa, legata

al valore economico del gioco. Termini come "buy", "refund" e "waste" indicano che molti utenti percepiscono l'acquisto del gioco come uno spreco di denaro e richiedono persino rimborsi. Altri termini come "garbage", "worth", e "worst" rinforzano questa narrativa di insoddisfazione. Questi risultati indicano una critica fortemente legata al rapporto qualità-prezzo. Gli utenti non solo percepiscono il gioco come insoddisfacente, ma sembrano sentirsi ingannati, fino al punto di considerarlo "spazzatura" e di richiedere un rimborso.

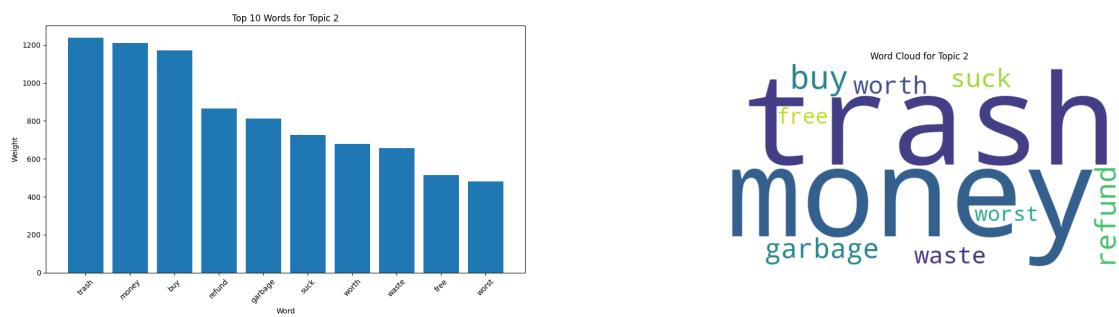


Figura 3.28: Terzo Topic Battlefield 2042

Dall'osservazione del grafico a barre e della word cloud 3.29, emergono alcune parole chiave che riflettono le opinioni degli utenti. Tra queste, spiccano termini come "good", "fun", "friend", "really" e "pretty", che suggeriscono come l'esperienza di gioco sia generalmente considerata piacevole e divertente. Molti giocatori sembrano apprezzare non solo la qualità del gameplay, ma anche l'aspetto sociale del titolo, dato che la parola "friend" evidenzia quanto sia comune condividere questa esperienza con amici.

Accanto a queste impressioni positive, però, compaiono anche termini più critici come "toxic" e "trash", che lasciano intendere la presenza di alcune problematiche. Inoltre, la presenza del termine "1010", probabilmente riferito a una valutazione massima ("10/10"), e la parola "buy" suggeriscono che molti utenti consigliano caldamente l'acquisto del gioco, riconoscendone il valore.

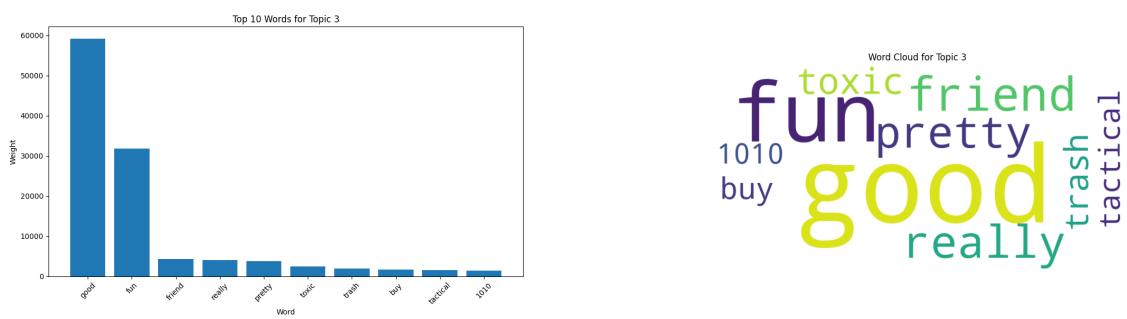


Figura 3.29: Quarto Topic Rainbow Six Siege

CAPITOLO 4

Conclusioni

In conclusione, dall'analisi congiunta dei topic, del sentiment e dell'emotio legate alle recensioni di Battlefield 2042 e Rainbow Six Siege emergono differenze significative nella percezione del pubblico.

Per quanto riguarda Battlefield 2042, i risultati del topic modeling evidenziano alcuni aspetti positivi, come un gameplay divertente e un certo interesse per i futuri aggiornamenti del gioco. Tuttavia, queste qualità positive sono fortemente oscurate da numerosi problemi segnalati dagli utenti, tra cui server instabili, mappe poco apprezzate, bug persistenti, gestione dei veicoli e controlli delle armi poco soddisfacenti. A ciò si aggiunge la percezione di un prezzo di acquisto troppo elevato rispetto alla qualità offerta. L'analisi del sentiment conferma una generale insoddisfazione nei confronti del gioco, mentre l'analisi delle emozioni mostra una presenza limitata di emozioni positive. Nonostante ciò, si registra un leggero miglioramento dell'opinione pubblica nel tempo, sebbene la percezione generale rimanga mediocre.

Al contrario, Rainbow Six Siege presenta un quadro più favorevole. I risultati del topic modeling mettono in luce diversi aspetti positivi, tra cui il divertimento derivante dalla componente collaborativa, specialmente quando giocato con amici, e un prezzo percepito come adeguato. Tuttavia, non mancano critiche rivolte a specifici aspetti del gioco, come alcuni operatori poco apprezzati, un sistema di shooting

ritenuto poco coinvolgente e la presenza di una community considerata tossica. L'analisi del sentiment riflette un apprezzamento generale più elevato rispetto a Battlefield 2042, mentre l'analisi delle emozioni evidenzia una predominanza di emozioni positive. È interessante notare come l'opinione su Rainbow Six Siege abbia raggiunto un picco positivo nel 2020, per poi calare progressivamente nel tempo.

In sintesi, Rainbow Six Siege gode di una reputazione complessivamente più positiva rispetto a Battlefield 2042, grazie a una maggiore soddisfazione generale e a emozioni più favorevoli espresse dagli utenti. Tuttavia, entrambi i titoli mostrano aree di miglioramento che potrebbero essere affrontate per ottimizzare l'esperienza di gioco e rispondere meglio alle aspettative della community.

Bibliografia

- [1] ActivePlayer.io, "Steam player count and statistics (2023)," 2023, accessed: December 6, 2024. [Online]. Available: <https://activeplayer.io/steam-player-count-statistics/> (Citato alle pagine iii e 4)
- [2] T. Wijman, "Newzoo's review of 2023: An epic year of video game data," *Newzoo*, 2023. [Online]. Available: <https://newzoo.com/resources/blog/video-games-in-2023-the-year-in-numbers> (Citato a pagina 1)
- [3] C. Damiano, "Videogiochi, il settore vola a 196 miliardi di dollari: vale più di cinema e streaming," *Business People*, 2024. [Online]. Available: <https://www.businesspeople.it/business/economia/videogiochi-il-settore-vola-a-196-miliardi-di-dollari-vale-piu-di-cinema-e-streaming/> (Citato a pagina 1)
- [4] V. Corporation, "Steam statistics: Annual review 2024," 2024, statistiche ufficiali pubblicate da Valve Corporation. [Online]. Available: <https://store.steampowered.com/> (Citato a pagina 3)