



International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Steam Game Review Sentiment Analyzer: A Review Analysing System Using NLP and Visualization

Gautam Suraj¹, K. S. Vishesh², Aryan Dewangan³, Monika Vishwakarma⁴

Computer Science and Engineering Department, Shri Shankaracharya Technical Campus, Bhilai, Chhattisgarh, India
gtm.suraj20@gmail.com¹; ksvishesh31@gmail.com²; aryandewangan881@gmail.com³; msvishofficial@gmail.com⁴

ABSTRACT

This paper introduces the design and development of a comprehensive Steam Game Review Sentiment Analyzer—an intelligent system built to automate the collection, classification, and visualization of public opinions expressed on Steam. By combining Natural Language Processing (NLP), sentiment analysis, API-based data collection, and modern visualization techniques, the system analyses up to 100 recent game reviews per title in real time.

Reviews are collected via Steam's public API, cleaned using NLTK, analysed for sentiment polarity using the VADER model, and stored in game-specific SQLite tables. A user-facing Streamlit interface allows users to input game URLs, select date ranges, and instantly view dynamic charts, word clouds, and sentiment distributions. Review data is deduplicated for accuracy and preserved across sessions in both a local database and a persistent CSV file. An embedded debug log also supports transparency and user-level diagnostics.

This system provides a complete, modular, and user-friendly solution for researchers, developers, and analysts seeking insight into player sentiment and behaviour across the Steam ecosystem.

Keywords: *Sentiment Analysis, Steam Reviews, Natural Language Processing, VADER, Streamlit, SQLite, Data Visualization, Review Scraping.*

I. INTRODUCTION

In recent years, the gaming industry has witnessed explosive growth, evolving into one of the most dominant and influential entertainment sectors globally. With millions of active users, online gaming platforms such as Steam have transformed how games are distributed, reviewed, and consumed. Steam, in particular, serves as a digital marketplace and social platform where players not only purchase and play games but also leave feedback in the form of user reviews. These reviews, which reflect real-time user experiences, carry substantial value—both for consumers seeking buying guidance and for developers striving to understand audience reception.

However, the vast quantity and unstructured nature of this review data pose a significant challenge. While individual reviews may offer insights into gameplay quality, bugs, balance, or story depth, it is virtually impossible to analyse thousands of such entries manually. This is where Natural Language Processing (NLP) and sentiment analysis come into play. These technologies allow us to automate the understanding of human language and identify subjective tones such as positivity, negativity, or neutrality within large text corpora.

The Steam Game Review Sentiment Analyzer is a full-stack, automated system designed to bridge the gap between raw user reviews and actionable insights. The system empowers users to enter any Steam game URL, fetch reviews through the Steam API, and analyse them using NLP models—all within an easy-to-use web interface. It performs data cleaning, sentiment scoring, and filtering to deliver a clear understanding of how a game is being received by its audience.

The platform not only offers sentiment classification but also supports advanced filtering features such as custom date ranges, helping users pinpoint when sentiment shifted or spiked. With visualizations like line charts, pie graphs, scatter plots, and word clouds, users gain access to a wide array of analytical tools. A built-in logging feature even allows users to trace data processing steps in real time, making it easier to understand or debug the system's behaviour.

What sets this system apart is its complete automation, lightweight deployment, and visual storytelling. There is no need for external scraping tools or browser automation; all reviews are gathered directly via structured API calls, ensuring reliability and performance. Furthermore, review data is stored locally using SQLite and exported as CSVs, offering both modular access and long-term persistence.

This tool is not only valuable for developers and publishers seeking feedback on their games but also serves researchers, marketers, and academics who are interested in exploring patterns in digital content consumption. By combining modern data science tools with a user-first approach, this project demonstrates how sentiment analysis can uncover the voice of the gaming community in an accessible, scalable, and insightful manner.

II. LITERATURE REVIEW

Sentiment analysis has been widely explored across digital platforms like Amazon and Twitter, where vast amounts of opinionated user-generated content offer fertile ground for analysis. These platforms typically feature brief and informal textual data, making them ideal for both lexicon-based approaches and machine learning classifiers. Over time, research has demonstrated the power of sentiment analysis in tracking public opinion, understanding consumer behaviour, and detecting feedback trends in real time.

Steam, as a gaming platform, provides a unique kind of dataset for sentiment research. Unlike traditional review sources, Steam reviews often include detailed narratives and metadata such as hours played, date posted, and recommendation status. This combination of structured and unstructured data allows for deeper behavioural insights into player experiences—especially when analysed over time or across multiple titles.

The VADER sentiment analysis model has become increasingly popular for such tasks due to its sensitivity to informal text, abbreviations, and emoticons—common in user-generated reviews. Its ability to provide compound polarity scores makes it lightweight, interpretable, and ideal for real-time web applications. Similarly, Streamlit has empowered data scientists to build intuitive dashboards, and SQLite offers simplicity in managing modular, game-specific data.

Recent developments in public APIs have also improved the accessibility and stability of web data collection. This has allowed systems like ours to abandon traditional browser automation in favor of direct API integration, resulting in faster, more reliable, and more scalable solutions. While prior research has tackled sentiment classification or review visualization independently, our work combines all these modern tools into a unified analytical engine specifically designed for the gaming domain.

III. METHODOLOGY

The system architecture is organized into five modular stages:

a. Frontend Input Interface

Made using Streamlit, users input a Steam game URL via a simple web form. A regular expression extracts the numeric game ID from the link, which is then passed to the backend for review collection and analysis.

b. Backend Review Scraper

The backend constructs API requests to the official Steam review endpoint, retrieving structured JSON data containing reviews, timestamps, playtime, and recommendation status. The system collects up to 100 reviews, balancing depth with performance. Duplicate entries are removed based on review text and date to maintain dataset integrity.

c. NLP and Sentiment Classification

The review text is cleaned using NLTK processes, including stopword removal and tokenization. Sentiment scoring is handled by the VADER SentimentIntensityAnalyzer, which computes a compound score and classifies each review as Positive, Neutral, or Negative.

d. Database and CSV Storage

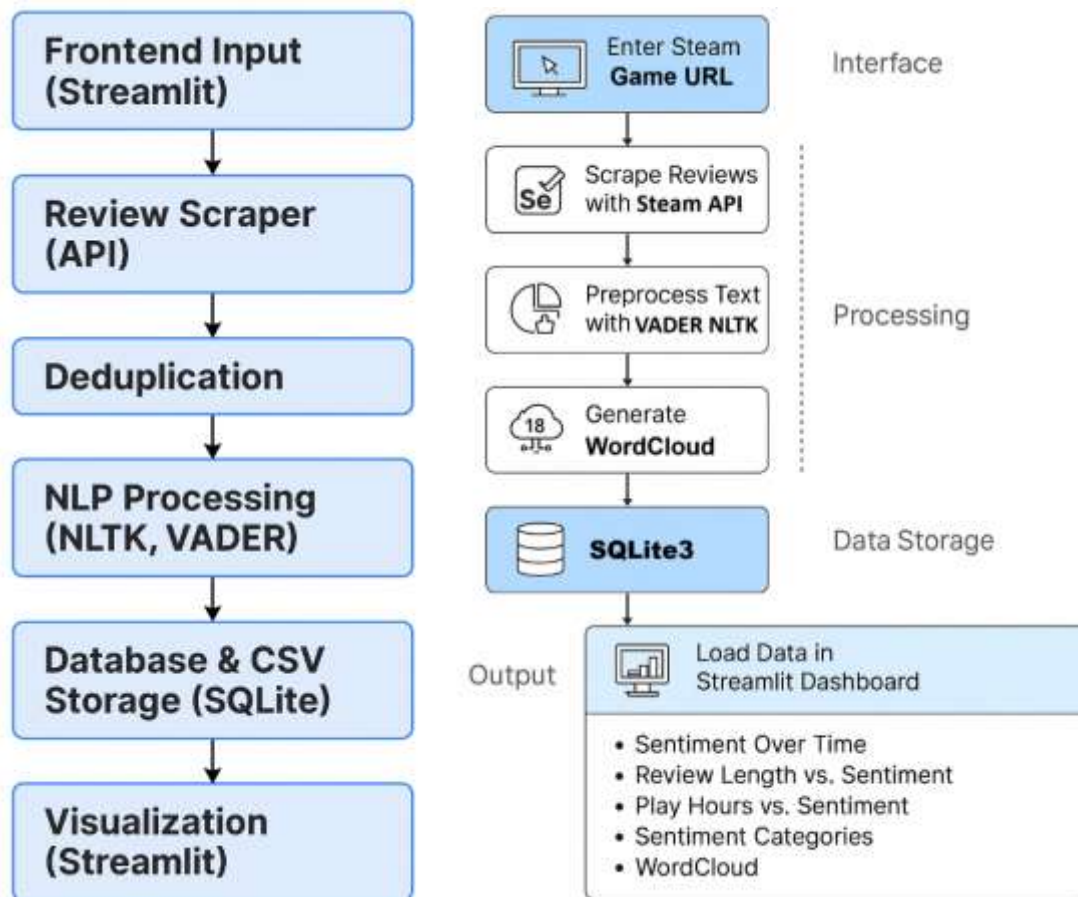
Cleaned reviews are stored in a game-specific SQLite table and also logged in a unified CSV file (game_reviews.csv) for session-spanning persistence. Deduplication logic is applied at both levels to prevent redundancy.

e. Interactive Visualization Dashboard

Once processing is complete, a Streamlit-based dashboard opens automatically. Users can filter reviews by custom date ranges and explore visualizations such as:

1. Sentiment trends over time
2. Playtime vs. sentiment scatter plots
3. Review length vs. polarity
4. Sentiment distribution pie charts
5. WordCloud of most-used terms

The dashboard also includes a live debug log panel to track date parsing, filtering logic, and potential issues during review loading. Figures below show the workflow and processing flowchart of the program.



IV. RESULT

Upon entering a valid Steam game URL, the system generates detailed, filterable visual outputs:

a. Sentiment Distribution

Reviews are classified into three sentiment categories, with real-time feedback on the distribution.

b. Time Trend Graphs

Average sentiment values are plotted over the selected date range.

b. Playtime & Length Analysis

Correlation patterns between sentiment, hours played, and review length are visualized.

Users with higher playtime often posted more detailed reviews, though sentiment varied.

c. WordCloud

Key themes in user feedback emerge through frequent word representation..

d. Debugging

A built-in log records potential issues like invalid dates or filter mismatches, aiding reproducibility and robustness.



Steam Game Review Scraper

Enter Steam Game URL:

<https://store.steampowered.com/app/105600/Terraria/>

URL

Start Date

2024/04/16

End Date

2025/04/16



← Press to start scrapping

Scraping reviews for Game ID 105600 between 16-04-2024 and 16-04-2025...



Logs

Found review on: 20-10-2024
Found review on: 23-08-2024
Found review on: 30-07-2024
Found review on: 03-07-2024
Found review on: 10-06-2024
Found review on: 23-05-2024
Found review on: 05-05-2024
Found review on: 04-05-2024
Found review on: 02-05-2024
Found review on: 30-04-2024
Found review on: 27-04-2024
Found review on: 14-09-2024
Found review on: 28-05-2024
Found review on: 08-07-2024

Reviews being saved

Done! Opening dashboard...

Select Game (ID)

search_582500

Filters

Select Date Range

2024/04/16 - 2025/04/16

Number of Top Reviews per Category

0 10 20

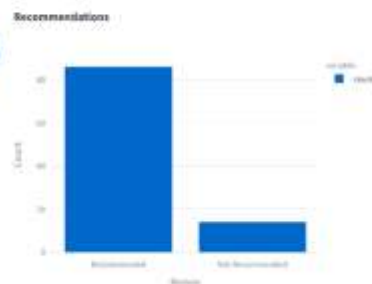
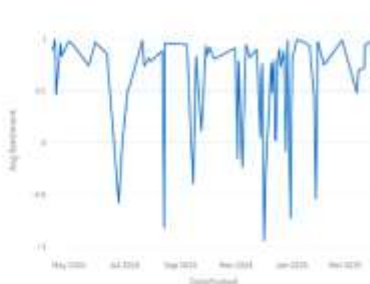
Export Filtered Data

Download Reviews as CSV

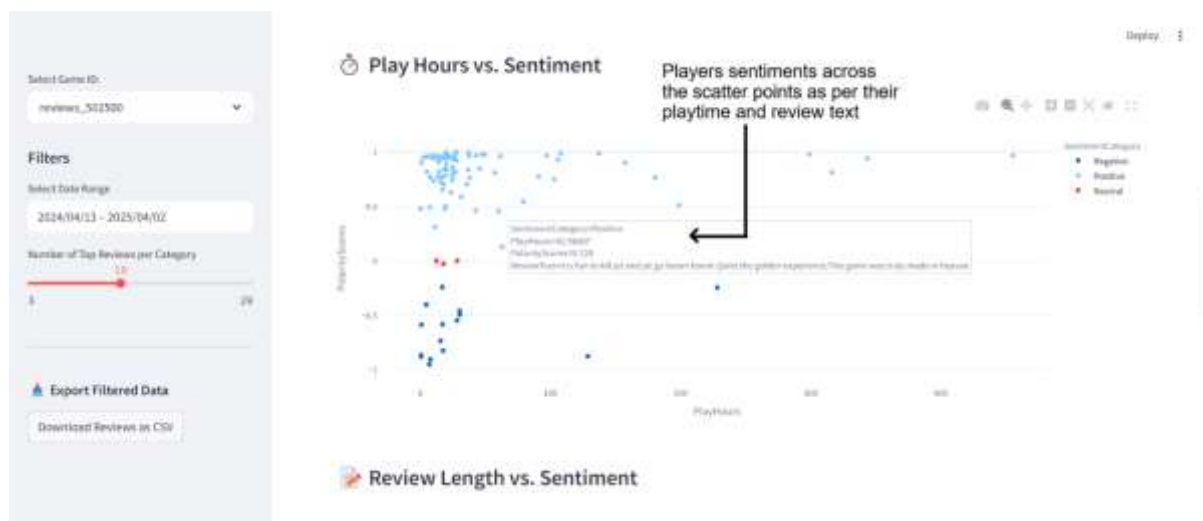
Steam Game Review Sentiment Analyzer

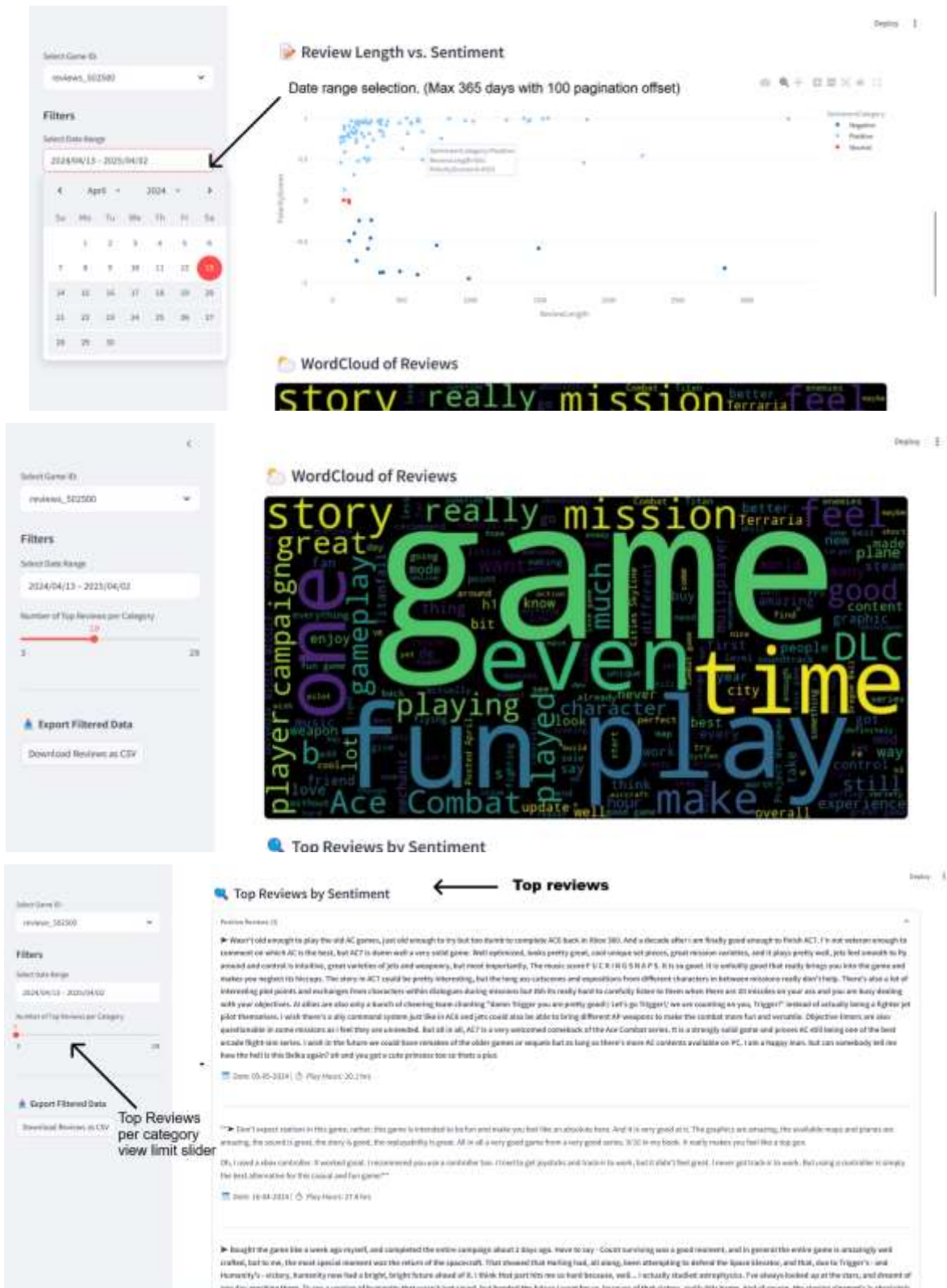
Sentiment Over Time

Recommendation Counts



Sentiment Distribution







V. CONCLUSION

The Steam Game Review Sentiment Analyzer provides a scalable, interactive, and user-friendly platform to analyze the sentiment landscape of PC games on Steam. By utilizing modern NLP and visualization frameworks, the system allows users to move from raw reviews to actionable insights within minutes.

Its modular design supports future enhancements such as cross-game sentiment comparison, multilingual review handling, or cloud-based hosting. With its current capabilities, the system already serves as a valuable tool for academics, game studios, and digital sociologists interested in understanding the narrative fabric of online gaming communities.

REFERENCES

- [1] C. J. Hutto and E. Gilbert, "VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text," ICWSM, 2014.
- [2] Streamlit Official Documentation – <https://docs.streamlit.io/>
- [3] Steam Web API – <https://partner.steamgames.com/doc/webapi>
- [4] NLTK Toolkit – <https://www.nltk.org/>
- [5] Steam Web Documentation : <https://steamcommunity.com/dev>
- [6] SQLite – <https://www.sqlite.org/index.html>
- [7] Plotly Express – <https://plotly.com/python/plotly-express/>
- [8] NLP Tutorial : https://www.tutorialspoint.com/natural_language_toolkit/index.html