IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 6, June 2025
Impact Factor: 7.67

Mining Insights from Sports Game Reviews with an Aspects based Sentiment Analysis Framework

Asst. Professor A. B. Bavane[1], Bhagat Gaurav Namdev[2], Patil Om Dilip[3], Gawade Prashant Dnyaneshwar[4], Kamathe Prachi Gangadhar[5]
Asst. Prof. Department of Information Technology [1]
Students, Department of Information Technology[2,3,4,5]
Dr. Vithalrao Vikhe Patil College of Engineering, Ahmednagar

Abstract: *Player-versus-player games and its accompanying tournament expansion have propelled esports games into a rapidly growing force in the gaming industry. Despite the strong cooperation between professional esports teams and operators, the voices of novice and armature players are sometimes disregarded due to a lack of useful analytical techniques. It is crucial to take into account the opinions of amateur players and carefully examine their reviews in order to guarantee the caliber of esports game services and create a fair gaming environment. This study suggests a novel approach for examining player evaluations in esports. Sentiment analysis and topic modeling are its two main components. Diverse subjects within reviews are efficiently identified by the framework through the use of the Latent Dirichlet Allocation (LDA) method. These themes were then used in a prevalence analysis to determine the relationships between players' worries and different esports games. Additionally, it makes use of a Transformer (TFM) downstream layer in conjunction with state-of-the-art Bidirectional Encoder Representations from Transformers (BERT) to accurately detect players' feelings toward various themes.*

Keywords*:* Esports, topic modeling, prevalence analysis, sentiment analysis, steam

I. INTRODUCTION

Esports has surged in popularity, captivating over two billion global players and spectators, and generating billions in annual revenue. Unlike traditional games, esports focuses on intense Player versus Player (PvP) competition, enhancing viewer engagement. With the rise of digital platforms like Steam, vast amounts of user-generated game reviews provide valuable insights but also pose challenges due to their unstructured nature. This paper presents a novel approach using aspect-based sentiment analysis (ABSA) to analyze and categorize esports game reviews, offering actionable insights for game developers and operators.

Motivation

The purpose of the sentiment analysis framework is to extract and analyze detailed feedback from game reviews to provide esports operators with actionable insights. By understanding players' sentiments towards various aspects of their games, operators can make informed decisions to enhance game quality and address player concerns effectively.

The purpose of the study is to develop a comprehensive framework that can analyzeesports game reviews from a broad range of players to identify key topics and sentiments, which can help game operators improve their services.

II. LITERATURE REVIEW

This paper [1]presents another broadly useful estimation dictionary Sentiment Lexicon and contrasts it and five existing vocabularies: Hu and Liu Opinion Lexicon, Multi-viewpoint Question Answering (MPQA) Subjectivity Lexicon, General Inquirer, National Research Council Canada (NRC) Word-Sentiment Association Lexicon and Semantic Orientation Calculator vocabulary. The adequacy of the opinion dictionaries for estimation order at the report level and

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 5, Issue 6, June 2025 Impact Factor: 7.67

sentence level was assessed utilizing an Amazon item survey informational collection and a news features informational index.

This paper [2] proposed Aspect based feeling grouping techniques are accessible in the writing however extremely restricted research work has focused on the programmed viewpoint recognizable proof and extraction of verifiable, rare, and coreferential angles. Viewpoint based characterization experiences the nearness of insignificant sentences in a regular client survey. Such sentences make the information uproarious and corrupt the characterization exactness of the AI calculations. This paper exhibits a fluffy angle based conclusion characterization framework which effectively removes viewpoints from client sentiments and perform close to precise grouping.

In this paper [3] Assessment mining or feeling examination is the computational investigation of individuals' sentiments, evaluations, frames of mind, and feelings toward substances, for example, items, administrations, associations, people, occasions, and their various angles. It has been a functioning exploration zone in regular language preparing and Web mining as of late. Scientists have contemplated assessment mining at the report, sentence and angle levels. Viewpoint level (called angle based assessment mining) is regularly wanted in functional applications as it gives the definite suppositions or estimations about various parts of elements and substances themselves, which are normally required for activity. Viewpoint extraction and substance extraction are in this way two center undertakings of angle based supposition mining.

This paper [4] presents a music recommendation framework dependent on an assessment force metric, named improved Sentiment Metric (eSM) that is the relationship of a vocabulary based estimation metric with a remedy factor dependent on the client's profile. This remedy factor is found by methods for abstract tests, led in a research center condition. In light of the test results, the remedy factor is defined and used to alter the last supposition force. The clients' assumptions are separated from sentences posted on interpersonal organizations and the music proposal framework is performed through a system of low multifaceted nature for cell phones, which recommends melodies dependent on the present client's slant force. Likewise, the structure was manufactured thinking about ergonomic criteria of ease of use.

This paper [5] presents an empirical comparison between SVM and ANN regarding document-level sentiment analysis. We discuss requirements, resulting models and contexts in which both approaches achieve better levels of classification accuracy. We adopt a standard evaluation context with popular supervised methods for feature selection and weighting in a traditional bag-of-words model. Except for some unbalanced data contexts, our experiments indicated that ANN produce superior or at least comparable results to SVM's. Especially on the benchmark dataset of Movies reviews, ANN outperformed SVM by a statistically significant difference, even on the context of unbalanced data.

In this study [6], they conduct a comparative assessment of the performance of three popular ensemble methods (Bagging, Boosting, and Random Subspace) based on five base learners (Naive Bayes, Maximum Entropy, Decision Tree, K Nearest Neighbor, and Support Vector Machine) for sentiment classification. Moreover, ten public sentiment analysis datasets were investigated to verify the effectiveness of ensemble learning for sentiment analysis. Based on a total of 1200 comparative group experiments, empirical results reveal that ensemble methods substantially improve the performance of individual base learners for sentiment classification.

This work [7]proposes an expansion of Bing Liu's viewpoint based feeling mining approach so as to apply it to the travel industry space. The expansion worries with the way that clients allude distinctively to various types of items when composing surveys on the Web. Since Liu's methodology is centered around physical item audits, it couldn't be straightforwardly applied to the travel industry area, which presents includes that are not considered by the model. Through an itemized investigation of on-line the travel industry item surveys, also found these highlights and afterward model them in our expansion, proposing the utilization of new and progressively complex NLP-based standards for the undertakings of abstract and supposition arrangement at the viewpoint level. additionally involve the undertaking of feeling perception and rundown and propose new strategies to assist clients with processing the tremendous accessibility of sentiments in a simple way.

In this paper [8], author propose a novel method to identify opinion features from online reviews by exploiting the difference in opinion feature statistics across two corpora, one domain-specific corpus (i.e., the given review corpus) and one domain-independent corpus (i.e., the contrasting corpus). We capture this disparity via a measure called domain relevance (DR), which characterizes the relevance of a term to a text collection. We first extract a list of

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 5, Issue 6, June 2025 Impact Factor: 7.67

candidate opinion features from the domain review corpus by defining a set of syntactic dependence rules. For each extracted candidate feature, we then estimate its intrinsic-domain relevance (IDR) and extrinsic-domainrelevance (EDR) scores on the domain-dependent and domain-independent corpora, respectively. Candidate features that are less generic (EDR score less than a threshold) and more domain-specific (IDR score greater than another threshold) are then confirmed as opinion features.

In this paper [9],author propose a sentiment classification method for the categorization of tourist reviews according to the sentiment expressed. We also give the results of the application of our sentiment analysis method on a real data set extracted from the AmFostAcolo tourist reviewWeb site. In our analysis we were focused on investigating the relation between the opinion holder and the accuracy of the review sentiment with the review score.

In this paper [10], author solve the problem in a different setting where the user provides some seed words for a few aspect categories and the model extracts and clusters aspect terms into categories simultaneously. This setting is important because categorizing aspects is a subjective task for different application purposes, different categorizations may be needed. Some form of user guidance is desired. In this paper, author propose two statistical models to solve this seeded problem, which aim to discover exactly what the user wants.

Existing System

The existing sentiment analysis system primarily focuses on basic preprocessing techniques such as tokenization, stopword removal, and stemming. After preprocessing, the data is directly fed into a classification model with minimal feature extraction. This often results in limited accuracy because the model may not effectively capture the semantic and contextual meanings of the text. The classification is generally rule-based or uses traditional machine learning algorithms with manually selected features. Consequently, this limits the system's ability to handle complex sentence structures or slang commonly found in user reviews, often leading to inaccurate sentiment predictions.

Proposed System

In the proposed system, a more advanced and structured pipeline is implemented for improved sentiment classification. The review dataset still undergoes essential preprocessing steps like stemming, stopword removal, and tokenization. However, the key enhancement lies in the **feature extraction** stage, where sophisticated techniques such as TF-IDF, word embeddings (like Word2Vec or BERT), or n-gram models are applied to better capture the contextual meaning of the text. These rich features are then input into a robust classification model, possibly using machine learning or deep learning algorithms, to improve accuracy. The final output classifies the sentiment as either positive or negative. This proposed model enhances the performance and reliability of the system in real-world applications where user reviews are diverse and often ambiguous

III. ARCHITECTURE

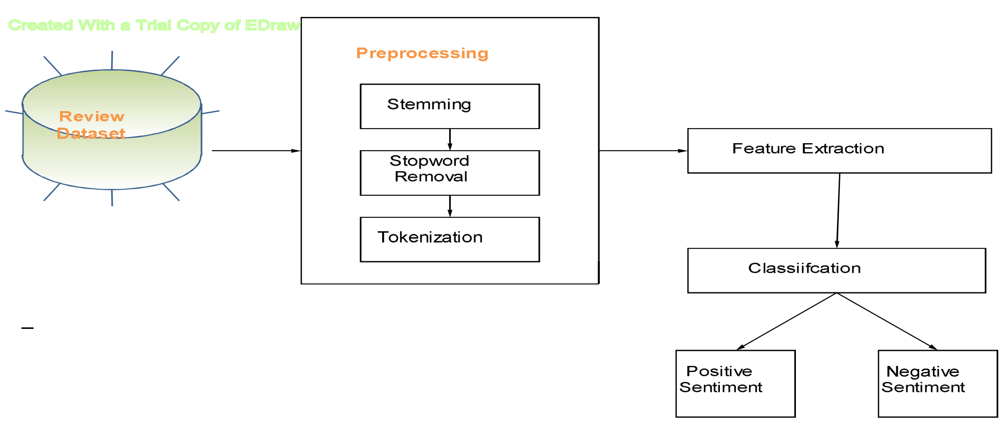

Fig : Proposed System Architecture

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 6, June 2025
Impact Factor: 7.67

IV. ALGORITHMS

Latent Dirichlet Allocation (LDA) Algorithm:

First and foremost, LDA provides a generative model that describes how the documents in a dataset were created. In this context, a dataset is a collection of D documents. Document is a collection of words. So our generative model describes how each document obtains its words. Initially, let's assume we know K topic distributions for our dataset, meaning K multinomial containing V elements each, where V is the number of terms in our corpus. Let βi represent the multinomial for the ith topic, where the size of βi is V: $|\beta i|=V$. Given these distributions, the LDA generative process is as follows:

Steps:

1. For each document:

(a) Randomly choose a distribution over topics (a multinomial of length K)

(b) for each word in the document:

(i) Probabilistically draw one of the K topics from the distribution over topics obtained in (a), say topic βj

(ii) Probabilistically draw one of the V words from βj

Random Forest

The algorithm used here is Random Forest. Random Forest is the most popular and powerful algorithm of machine learning.

Step 1: Assume N as number of training samples and M as number of variables within the classifier.

Step 2: The number m as input variables to decide the decision at each node of the tree; m should be much less than M.

Step 3: Consider training set by picking n times with replacement from all N available training samples. Use the remaining of the cases to estimate the error of the tree, by forecasting their classes.

Step 4: Randomly select m variables for each node on which to base the choice at that node. Evaluate the best split based on these m variables in the training set.

Step 5: Each tree is fully grown and not pruned (as may be done in constructing a normal tree classifier). For forecasting, a new sample is pushed down the tree. It is assigned the label of the training sample in the terminal node it ends up in.

V. RESULTS AND DISCUSSION

Parameter	Existing System	Proposed System
Precision	68.45	78.70
Recall	79.44	65.64
F-Measure	72.11	74.31
Accuracy	79.29	86.36

Stepwise Reviews Processing:

ID	Reviews	Stemming	Stopwords Removal	Algorithm Result
1	game is very nice. graphics are smooth .	game is very nice graphics ar smooth	game nice graphics ar smooth	Graphics
7	many server issue. worst game.	many server issu worst game	server issu worst game	Graphics
8	Im playing since season 8 It was awesome but time by time its getting worse even with full network always request timed out and its keep saying login again Sorry to say but pubg is going to lose players if its gonna be like that	im play sinc season 8 it was awesom but time by time its get wors even with full network always request time out and its keep say login again sorry to say but pubg is go to lose players if its gonna be like that	im play sinc season 8 awesom time time get wors even full network always request time keep say login sorry say pubg go lose players gonna	Server

VI. CONCLUSION

In esports, unbalanced gaming experience between professional players and noobs is a problem caused by the lack of methods for operators to quickly obtain useful information from noob players' feedback. To address this, we propose a hybrid approach of topic modeling and sentiment analysis to automatically analyze the vast number of game reviews from noob players. This will enable esports operators to better target their opinions and build a more balanced gaming environment

Future Scope

In future we will enhance this system to implement an android application and also on images.

REFERENCES

[1] Muhammad Afzaal, Muhammad Usman, Alvis Fong," Tourism Mobile App with Aspect-BasedSentiment Classification Framework for Tourist Reviews" IEEE Transactions on Consumer Electronics Vol: 65, May 2019.
[2] C. S. Khoo and S. B. Johnkhan, "Lexicon-based sentiment analysis: Comparative evaluation of six sentiment lexicons," Jour. Inform. Scien., vol. 44, no. 4, pp. 491-511, Aug. 2018, DOI: 10.1177/0165551517703514
[3] M. Afzaal, M. Usman, A. C. M. Fong, S. Fong, and Y. Zhuang, "Fuzzy Aspect Based Opinion Classification System for Mining Tourist Reviews," Adva. In Fuzzy Sys., vol. 2016, Oct. 2016, DOI:10.1155/2016/6965725
[4] X. Ma and E. Hovy, ''End-to-end sequence labeling via bi-directional LSTM-CNNs-CRF,'' in Proc. 54th Annu. Meeting Assoc. Comput. Linguistics, Long Papers, vol. 1. Berlin, Germany: Association for Computational Linguistics, 2016, pp. 1064–1074, doi: 10.18653/v1/P16-1101.
[5] Muhammad Afzaal, Muhammad Usman, Alvis Fong," Tourism Mobile App with Aspect-BasedSentiment Classification Framework for Tourist Reviews" IEEE Transactions on Consumer Electronics Vol: 65, May 2019.

[6] C. S. Khoo and S. B. Johnkhan, “Lexicon-based sentiment analysis: Comparative evaluation of six sentiment lexicons,” Jour. Inform. Scien., vol. 44, no. 4, pp. 491-511, Aug. 2018, DOI: 10.1177/0165551517703514

[7] M. Afzaal, M. Usman, A. C. M. Fong, S. Fong, and Y. Zhuang, “Fuzzy Aspect Based Opinion Classification System for Mining Tourist Reviews,”Adva. In Fuzzy Sys., vol. 2016, Oct. 2016, DOI:10.1155/2016/6965725

[8] L. Zhang and B. Liu, “Aspect and entity extraction for opinion mining,” in Data Mining Knowledge Discovery For Big Data, Berlin, Germany: Springer, 2014, pp. 1-40, DOI: 10.1007/978-3- 642-40837-3_1

[9] R. L. Rosa, D. Z. Rodriguez, and G. Bressan, “Music recommendation system based on user's sentiments extracted from social networks,” IEEE Trans. Consum. Electron., vol. 61, no. 3, pp. 359-367, Aug. 2015, DOI: 10.1109/TCE.2015.7298296

[10] R. Moraes, J. F. Valiati, and W. P. G. Neto, “Document-level sentiment classification: An empirical comparison between SVM and ANN,” Expert Sys. With Appli., vol. 40, no. 2, pp. 621-633, Feb. 2013, DOI: 10.1016/j.eswa.2012.07.059

[11] G. Wang, J. Sun, J. Ma, K. Xu, and J. Gu, “Sentiment classification: The contribution of ensemble learning,” Decision Supp. Sys., vol. 57, pp. 77-93, Jan. 2014, DOI: 10.1016/j.dss.2013.08.002

[12] E. Marrese-Taylor, J. D. Velásquez, and F. Bravo-Marquez, “A novel deterministic approach for aspect-based opinion mining in tourism products reviews,” Expert Sys. With Appli., vol. 41, no. 17, pp. 7764-7775, Dec. 2014, DOI: 10.1016/j.eswa.2014.05.045

[13] Z. Hai, K. Chang, J.-J. Kim, and C. C. Yang, “Identifying features in opinion mining via intrinsic and extrinsic domain relevance,” IEEE Trans. Know. And Data Engi., vol. 26, no. 3, pp. 623-634, Mar. 2014, DOI: 10.1109/TKDE.2013.26

[14] M. Colhon, C. Bădică, and A. Şendre, "Relating the opinion holder and the review accuracy in sentiment analysis of tourist reviews," in Int. Conf. Knowledge Sci., Eng. and Manage., 2014, pp. 246-257, DOI: 10.1007/978-3-319-12096-6_22

[15] A. Mukherjee and B. Liu, "Aspect extraction through semisupervised modeling," in Proc. 50th Annu. Meeting Assoc. for Computational Linguistics, 2012, pp. 339-348