National College of Ireland

Project Submission Sheet

Student Name:	Aryaman Chettri		
Student ID:	22135511		
Programme:	MSc Data Analytics	**Year:**	2025 - 2026
Module:	Research Practicum/ Master Thesis		
Lecturer:	Dr. Bharat Agarwal		
Submission Due Date:	28-01-2026		
Project Title:	Game Sentiment Prediction Using Deep Learning Regression		
Word Count:	7602		

Academic Integrity Certificate Number: adAXG3ZhHv

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the references section. Students are encouraged to use the Harvard Referencing Standard supplied by the Library. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action. Students may be required to undergo a viva (oral examination) if there is suspicion about the validity of their submitted work.

Signature:	Aryaman Chettri
Date:	28-01-2026

PLEASE READ THE FOLLOWING INSTRUCTIONS:

1. Please attach a completed copy of this sheet to each project (including multiple copies).
2. Projects should be submitted to your Programme Coordinator.
3. **You must ensure that you retain a HARD COPY of ALL projects**, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer. Please do not bind projects or place in covers unless specifically requested.
4. You must ensure that all projects are submitted to your Programme Coordinator on or before the required submission date. **Late submissions will incur penalties.**
5. All projects must be submitted and passed in order to successfully complete the year. **Any project/assignment not submitted will be marked as a fail.**
6. **Please check that you read AI and Academic Integrity Acknowledgement Supplments in this document**

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

AI Acknowledgement Supplement

Research Practicum/ Master Thesis

Game Sentiment Prediction Using Deep Learning Regression

Your Name/Student Number	Course	Date
Aryaman Chettri	MSc Data Analytics	28-01-2026

This section is a supplement to the main assignment, to be used if AI was used in any capacity in the creation of your assignment; if you have queries about how to do this, please contact your lecturer. For an example of how to fill these sections out, please click here.

AI Acknowledgment

This section acknowledges the AI tools that were utilised in the process of completing this assignment.

Tool Name	Brief Description	Link to tool
NA	NA	NA

Description of AI Usage

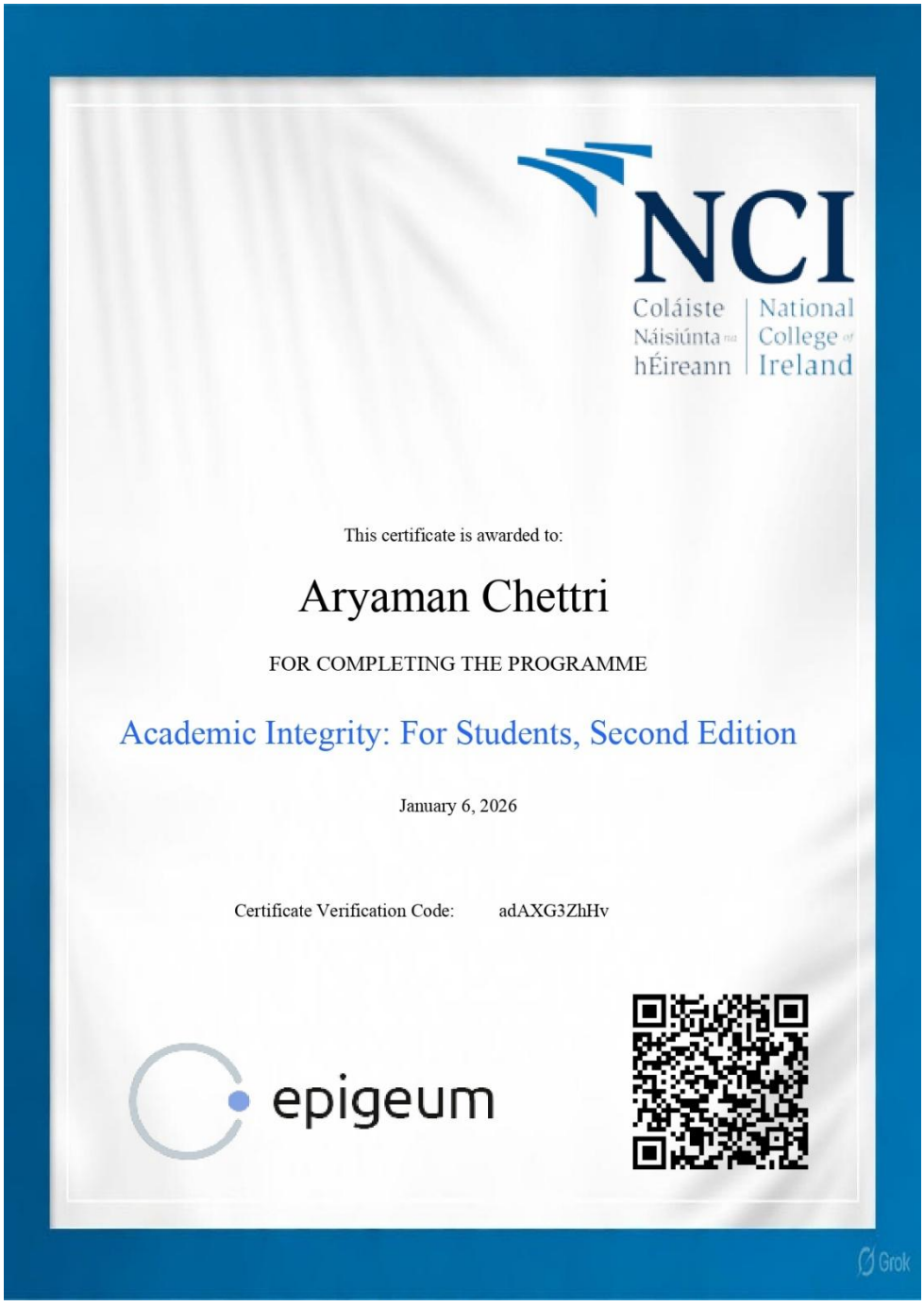
This section provides a more detailed description of how the AI tools were used in the assignment. It includes information about the prompts given to the AI tool, the responses received, and how these responses were utilised or modified in the assignment. **One table should be used for each tool used**.

[Insert Tool Name]	
NA	
NA	NA

Evidence of AI Usage

This section includes evidence of significant prompts and responses used or generated through the AI tool. It should provide a clear understanding of the extent to which the AI tool was used in the assignment. Evidence may be attached via screenshots or text.

Academic Integrity Acknowledgement Supplement



Academic Integrity Subject Guide or how to access the site:
https://libguides.ncirl.ie/academicintegrity/epigeumforstudents

Completion of Epigeum Academic Integrity Online Course is Mandatory for all registered students of NCI

National College of Ireland

Game Sentiment Prediction Using Deep Learning Regression

MSc Research Project
MSc in Data Analytics

Aryaman Chettri
Student ID: 22135511

School of Computing
National College of Ireland

Supervisor: Dr. Bharat Agarwal

National College of Ireland

MSc Project Submission Sheet

School of Computing

Student Name: ……Aryaman Chettri……

Student ID: 22135511……

Programme: MSc in Data Analytics…… **Year:** 2025 - 2026…..

Module: Research Practicum/ Master Thesis……

Supervisor: Dr. Bharat Agarwal……

Submission Due Date: 28-01-2026……

Project Title: Game Sentiment Prediction Using Deep Learning Regression……

Word Count: 7602…… **Page Count**…27……

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.
ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature: Aryaman Chettri……

Date: 28-01-2026……

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	□
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	□
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	□

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Game Sentiment Prediction Using Deep Learning Regression

Aryaman Chettri
22135511

Abstract

Gaming has evolved from paying for a physical game to monetisation ecosystems comprising microtransactions, battle passes, loot boxes, and gacha mechanics, and many of these systems have been publicly criticised for exploiting players. Steam reviews and Reddit threads are usual spots for feedback on games and game developers. Comparing the performance of deep learning regression models for predicting video games' sentiment scores based on independent game metadata features. Data included Reddit (posts, upvotes, comments), price, platform and downloadable content. Data for 1,413 games were retrieved using the Steam Web API and the Reddit API. 27 GDPR-compliant features were devised. The four architectures tested include LSTM, CNN, DNN and the hybrid model. Ground truth sentiment scores have been computed using VADER for Steam reviews. LSTM's MAE was 0.130, R^2 score 0.043, with a (-0.5, 0.8) range. This was 2.2% better than baseline and achieved the highest correlation among the 5 models (0.174 between features), indicating that the features were independent of one another. The resulting positive R^2 of 4.3% suggests that there is a useful signal in the metadata and a baseline performance suitable for production, with low sentiment scores correlated with excessive downloadable content, predatory free-to-play models, and controversy.

1 Introduction

1.1 Background

The gaming industry as a whole has seen remarkable growth over the past decades, and it has evolved into a global entertainment market worth billions of euros annually (approximately 206.5 billion EUR). This growth has also brought a major shift in how games generate revenue. What used to be a traditional model of purchasing complete games for a fixed price has now given way to an ecosystem of continuous monetisation from players through mechanics such as microtransactions, season passes, battle passes, loot boxes, and gacha systems, to name a few.

While these monetisation strategies have proven very profitable for the gaming industry, they have also raised major ethical concerns among gamers. Loot boxes, in some cases, have been compared to gambling and have faced multiple instances of scrutiny from the community and jurisdictions. The psychology manipulation techniques used encourage spending, the fear of missing out (FOMO), artificial scarcity and variable reward schedules, which have been

criticised as predatory, when we consider that the target audience population include children and individuals with gambling issues.
Community sentiments serve as a powerful and authentic proxy for verifying ethical concerns in game monetisation. When players detect monetisation concerns, the community's response is swift, vocal and measurable. Game's Steam review shifting from positive to negative. A whole Reddit thread of players' complaints about the game. This kind of community feedback creates rich data sources that can be used to identify unethical issues on a scale.

Deep learning techniques have made remarkable success across diverse prediction tasks, ranging from image recognition to natural language processing to time series forecasting. Recurrent neural networks, particularly the Long Short-Term Memory(LSTM) model, have shown great success in learning complex patterns from sequential and tabular data. This research explores whether such a technique can extract meaningful signals from game metadata to reveal community sentiment, providing a foundation for ethical analysis of game monetisation practices.

1.2 Motivation and Problem Statement

The motivation for this research was to develop a quantitative framework for evaluating ethical practices in game monetisation. During the initial research, the focus was on constructing an "ethical score" to measure the ethical level of the given game. However, this approach revealed a critical flaw that could have compromised the validity of the resulting models.

The target variable "ethics_score" was calculated using a "formula" that directly incorporated sentiment as a component:

$$\text{ethics\_score} = 50 + (\text{avg\_sentiment\_score} * 25) + \text{bonuses/penalties}$$

This formulation created a circular dependency in which the required variables were mathematically derived from features that were also used as predictors. The end result is catastrophic: R^2 values range from -108 to -0.017 for models trained using this formulation. These negative R^2 values indicated there was something fundamentally wrong with the performance of the models, requiring us to adapt and think of a newer approach.

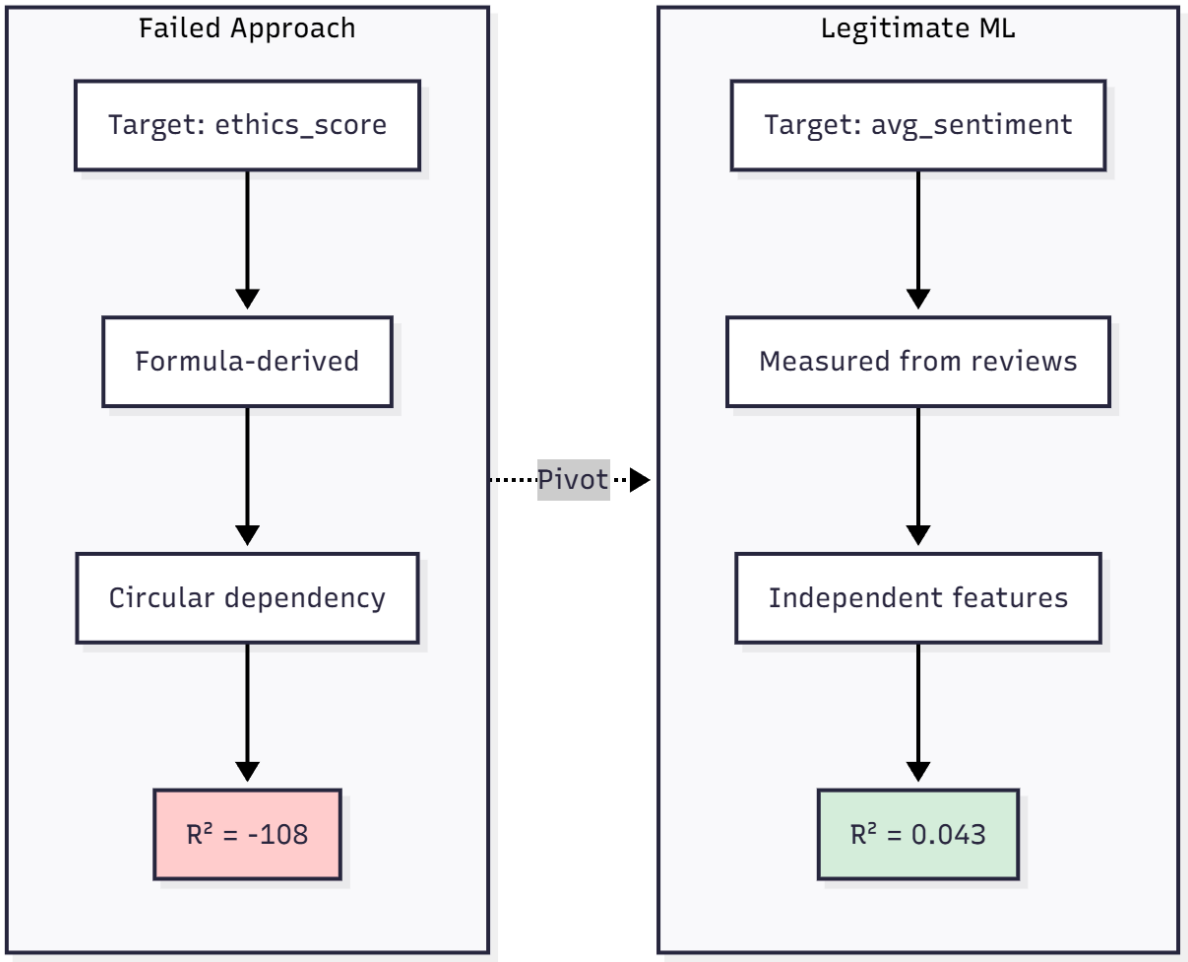


Figure 1: The Problem - Pivot from Circular Dependency to Legitimate ML

Figure 1 illustrates the changes made in this research. The failed approach (left) used the formula-derived "ethics_score," which created a circular dependency, resulting in negative R^2 values. The better approach (right) measured sentiment from independent features, achieved a positive R^2 value, a weak but genuine result that shows the validity of the method.

This highlights in machine learning the importance of genuine independence between input features and target variables. A model might achieve high accuracy by exploiting circular dependencies, providing no predictive value, and in the worst case, will fail when used on new data.

1.3 Research Question and Objectives

Can deep learning regression models predict game sentiment score from independent metadata features, and to what extent can such predictions serve to identify ethical concerns in game monetisation practices?

The research question helps us address the technical challenge of building a predictive model with verified feature independence and using sentiment prediction as a foundation for ethics analysis. To address the research question, four objectives were made:

1. Research the current state of the art in game sentiment analysis and ethics research, identify issues such as circular dependencies and make a foundation for ethics evaluation using sentiment as a proxy.
2. To design and implement a feature engineering pipeline that creates independent metadata features without direct mathematical or conceptual relationship to sentiment scores, validated through rigorous correlation analysis.
3. Implement and train multiple deep learning regression models, including LSTM, CNN, DNN, and Hybrid architectures, to predict sentiment from the engineered independent features.
4. Checking the model's performance using the appropriate regression metrics (MAE, RMSE, R^2) and validating the feature independence.

1.4 Contribution

Firstly, this research provides a pipeline for creating 27 independent features from game metadata that avoid circular dependencies with sentiment targets, with a correlation of only 0.174. This research can also be applied to other fields where feature independence is important, not just gaming. Second, providing a comparative evaluation of four deep learning models (LSTM, CNN, DNN, and Hybrid) for game sentiment regression. The results of LSTM provided better performance for the given task. Third, it is a ready to be used LSTM base model with genuine learning (positive R^2), which can be used as a tool for game sentiment evaluation. Fourth, the research shows how the analysis of community sentiment correlates with ethical concerns, negative sentiment often associated with aggressive free-to-play monetisation, DLC counts and controversy surrounding the game. Finally, an intuitive, colour-coded visual dashboard for interpreting and acting on model predictions.

1.5 Limitations

Several limitations hinder the current research and should be acknowledged. The dataset consists of 1,413 games, which is sufficient for showing the proof of concept; however, with a limited statistical power, it may not be able to capture the full diversity of the gaming market. Feature engineering is restricted to metadata independent of sentiment, which limits predictive power but ensures methodological validity.

We know players' opinions change over time. The research employs a snapshot approach, we collected the necessary data from a single point in time rather than tracking sentiment evolution over time. This means dynamic changes to sentiment following game updates, DLC releases or monetisation policy changes are not captured. Currently, the model cannot directly detect some monetisation mechanics, such as loot boxes or gacha systems, without additional natural language processing of review text, an opportunity for future enhancement for the research.

2 Related Work

This chapter looks at the existing research relevant to our work. We've organised the discussion into four connected themes: how the gaming industry makes money and how players behave, the methods used for sentiment analysis and how they've developed over

time, deep learning architectures for analysing text, and where sentiment analysis meets gaming. This structure helps us build a solid theoretical foundation whilst highlighting the gap in research that our study aims to fill.

2.1 Gaming Industry Monetisation and Player Behaviour

The gaming industry has gone through massive changes in how it makes money, gone are the days of the one-time purchase model. It has evolved into more complex systems that include microtransactions, loot boxes, and subscription services. Alić and Čić (Alić, 2024) carried out research looking at what users actually think about in-game monetisation tactics, particularly microtransactions and how engaged players are in CS:GO. What they found was that people really don't like pay-to-win mechanics where spending money gives you an unfair advantage in the game. This psychological aspect has a measurable effect on user reviews and how sentiment gets expressed.

Building on this, An et al. (An, 2024) looked into how buying game-related merchandise differs from making purchases within the game itself, using a mixed-method approach. Their analysis showed a link between how fair players think the monetisation system is and whether they stick around long-term. Importantly, they highlighted that the sentiment players express in reviews often predicts how they'll engage with the game later, suggesting that sentiment analysis could help predict business outcomes.

The framework put forward by Çetiner and Çalışkan (Çetiner, 2025) took this research further by examining what motivates Pokemon GO players to make in-game purchases through microtransactions. Their work showed that the distribution of positive and negative sentiment in gaming reviews has unique characteristics that are quite different from general product reviews, which means you need specialised approaches to analyse them properly. Additionally, Chaudhary and DaSouza (Chaudhary, 2024) explored how ready consumers are for microtransactions in digital content business models, showing that sentiment patterns shift in response to game updates, balance patches, and community events.

The psychological side of gaming engagement has attracted significant academic interest. Barros et al. (Barros, Veiga, Moura, Urdan, & Matos, 2024) In their research, they found the behavioural effects of how Steam Community Market users perceive value. They identified different sentiment patterns linked to different types of players. Competitive players tend to focus on balance issues, whilst casual players care more about accessibility and whether the game's actually fun. Lee et al. (Lee, Lee, Bowman, Yao, & Chen, 2024) added to this understanding by translating and validating the Video Game Demand Scale into Korean, finding that gaming demand and the social dynamics in multiplayer environments really do affect review sentiment.

Liu et al. (Liu, You, Zou, & Cao, 2024) provided crucial insights into modelling customer preferences for mobile games based on online reviews, using the BW-CNN and S-Kano models. Revealling that cultural backgrounds and what customers wants and expectation have a substantial influence on both the content and intensity of sentiment in game reviews. Surjandy and Cassandra (Surjandy, 2024) backed this up by investigating how gaming motivations affect gamers' willingness to invest, which laid the groundwork for sentiment analysis approaches that are adapted to the gaming domain.

2.2 Sentiment Analysis Methodologies

Sentiment analysis (also known as opinion mining) is defined as the extraction of subjective information from text. Initially, it was done using lexicon-based methods, but it has since evolved towards methods based on deep learning. C.B et al. (C.B, et al., 2025) used the lexicon-based VADER (Valence Aware Dictionary and sEntiment Reasoner) and TextBlob approaches to predict sentiment around social media content in real-time. Despite their efficiency, the research concludes that these lexicon-based approaches are not context-aware and thus can make them less effective in a narrow application such as gaming.

VADER (Valence Aware Dictionary and sEntiment Reasoner) is a social media oriented sentiment analysis tool that uses a lexicon and rules with grammatical heuristics and sentiment-oriented emoticons to compute the compound polarity score. S and Fernandez (S & Fernandez, 2025) implemented VADER and Support Vector Machine (SVM) classification as a hybrid model to provide real time classification and sentiment analysis of news articles. Their hybrid approach outperformed VADER alone, leveraging VADER's rule-based approach and machine learning based classifier to overcome VADER's pitfalls to good effect, and had good accuracy in education, finance, and health domains.

Comparative studies of sentiment analysis methods have been done to help choose between the methods. Thakur et al. (Thakur, et al., 2024) compared VADER, RoBERTa and DistilBERT to predict consumer sentiments, with NLP and pre-trained transformer-based models, on Amazon reviews. While VADER is great at quickly getting sentiment scores, the transformer models (like RoBERTa) do a much better job with subtle sentiments and dependencies. They suggest you choose models based on what you value more: accuracy, computational resources, turn-around time, or a speed-and-accuracy trade-off that works best for you.

With the progress in technology, machine learning and deep learning have provided better results in the prediction of sentiment compared to customary machine learning. Aamir, J and Sweety (Aamir, J, & Sweety, 2024) compare the performances of ML and DL approaches for sentiment analysis of tweets using the machine learning classifiers Support Vector Machine (SVM), Random Forest, Logistic Regression, and XGBoost, and deep learning algorithms Convolutional Neural Network and Feedforward Neural Network. The authors concluded that while interpretable baselines are sufficient, deep learning models are more accurate when the data is informal and casual and has high contextual complexity.

2.3 Deep Learning Architectures for Text Analysis

Deep learning has transformed natural language processing with its architectures that learn feature hierarchies from data. LSTM can learn long-term temporal dependencies. LSTM solves the vanishing gradient problem in customary recurrent neural networks through the use of gating to control the flow of information (Mienye & Swart, 2024). In Vu et al. (Vu, Pham, Solanki, Hoang, & Tran, 2024) evaluated the performance of machine learning and deep learning models with LSTM architecture and embedding layers against Naive Bayes and Random Forest classifiers on the IMDb movie reviews and financial sentiment datasets for sentiment classification.

Bidirectional LSTM (BiLSTM) is an extension of the LSTM structure, in which the input features are processed in both directions, from left to right and right to left, thus gaining

information for both previous and next elements. Khoirunnisa and Pamungkas (Khoirunnisa & Pamungkas, 2025) compared the performance of LSTM and BiLSTM algorithms to classify user review sentiment of the Shopee app. The BiLSTM outperformed unidirectional LSTM with an accuracy of 91.3% and an AUC of 0.96 due to its ability to capture bidirectional context which is essential for sentiment analysis.

The attention mechanism, originally developed to address the shortcomings of fixed-length context vectors, has been found to work well for sentiment analysis applications. Sakib et al. (Sakib, Chowdhury, Hira, Choudhury, & Das, 2025) compared the performance of LSTM with attention mechanism and BERT for sentiment analysis of social media and e-commerce reviews from the TikTok and NYKAA e-commerce platforms. Their results also showed that BERT outperformed other approaches, achieving 84% accuracy on TikTok and 81% on NYKAA, while long short-term memory (LSTM) with attention achieved 81% and 79% on TikTok and NYKAA, respectively. The work showed that attention mechanisms enable models to focus on the parts of text that contain sentiment, improving interpretability.

Transformer architecture-based models were proposed to outperform the state-of-the-art in many NLP tasks. Bidirectional Encoder Representations from Transformers (BERT) performs self-attention thus considering all words at once. This solves the long-term dependency issue of RNNs (Riatma, et al., 2025). Gupta et al. (Gupta, Jha, Singh, Bharti, & R, 2024) benchmarked logistic regression, BERT and other classifiers on the Amazon Fine Food Reviews dataset for the task of sentiment analysis of e-commerce products. The authors fine-tuned the BERT language model for this task and achieved a state-of-the-art score of 98.44% accuracy, precision of 98.73% and recall of 99.25% compared to 89.50% for Random Forest and 86.15% for Support Vector Machines.

Yan and Cui (Yan & Cui, 2025) replicated this study and examined the performance of ML and DL models for fine-grained sentiment classification of movie reviews. The study compared the performance of SVM, Random Forest, Logistic Regression, LSTM, and CNN using the NLTK Movie Reviews dataset and concluded that DL models, especially LSTM and CNN models, outperformed ML models for the task. Jonathan et al. (Jonathan, Halim, Wulandhari, & Nabiilah, 2025), amade use of a machine and deep learning approach to sentiment classification in conversations with ChatGPT, further proving that deep contextual models exceed shallow models in sentiment classification tasks.

2.4 Sentiment Analysis in Gaming Domains

Sentiment analysis in gaming applications also has some unique aspects to it. Pamuji, Fatichah and Navastara (Pamuji, Fatichah, & Navastara, 2024) investigated deep-learning-based topic modelling for gaming-related subjects (Apex Legends player reviews) using BERTopic and Top2Vec to identify topics that arise from the sentiment expressed in player reviews of the video game Apex Legends. Their work also illustrated how BERTopic using all-MiniLM-L6-v2 embeddings learned considerably better-coherent and more interpretable topics than customary topic models such as Latent Dirichlet Allocation (LDA) and Non-Negative Matrix Factorisation (NMF). As an example, their work discovered user dissatisfaction with technical problems and satisfaction with gameplay and character design.

The foundation for deep learning sentiment analysis has improved alot since the past years. Mienye and Swart (Mienye & Swart, 2024) provides a review on deep learning architectures,

advances, applications covering architectural considerations, training strategies and deployment methods. In addition, they also discuss the importance of feature engineering and embedding representation of the data, in order to achieve a satisfactory performance across different domains. In line with this, Riatma et al. (Riatma, et al., 2025) compare the prediction performance of deep learning architectures like LSTM and GRU when predicting product stock. Sivamayil et al. (Sivamayil, et al., 2023) systematically studied reinforcement learning-based applications and proposed new optimization and ensemble methods to improve classification robustness.

Preprocessing and feature extraction are important factors in the performance of sentiment classifiers. Preprocessing techniques include text normalisation, tokenisation, stopword removal, and lemmatisation. TF-IDF and word embeddings are widely used in converting the text into numerical format to learn models (Sakib, Chowdhury, Hira, Choudhury, & Das, 2025), such as using Word2Vec or GloVe. Contextual embeddings can largely remove the need for manual feature engineering and improve transfer learning between domains via the use of transformer architecture.

2.5 Research Gap and Contribution

Despite progress in game industry analysis and sentiment prediction, research at the intersection of the two is scarce. Previous work either focused on descriptive analysis of player sentiment or on general-domain sentiment classification, and neglected to examine the feasibility of predicting sentiment using game metadata as features. Additionally, deep learning regression approaches for predicting continuous sentiment scores have not been thoroughly explored in the context of gaming.

Table 1: Comparison of Sentiment Analysis Approaches in Literature

Approach	Study	Domain	Method	Performance	Limitation
Lexicon-based	C.B et al. (2025)	Social media	VADER, TextBlob	Fast, real-time	Not context-aware
Hybrid Lexicon+ML	S & Fernandez (2025)	News articles	VADER + SVM	Good accuracy	Domain-specific
Transformer-based	Gupta et al. (2024)	E-commerce	Fine-tuned BERT	98.44% accuracy	Computationally expensive
BiLSTM	Khoirunnisa & Pamungkas (2025)	E-commerce reviews	BiLSTM	91.3% accuracy	Classification only
LSTM + Attention	Sakib et al. (2025)	Social media	LSTM with attention	81% accuracy	Text-based input
Topic Modelling	Pamuji et al. (2024)	Gaming (Apex Legends)	BERTopic	Good coherence	Descriptive, not predictive

ML Comparison	Aamir et al. (2024)	Twitter	SVM, RF, CNN, FNN	DL outperforms ML	Informal text only
This Study	**—**	**Gaming (1,413 games)**	**LSTM regression**	**R^2 0.043, MAE 0.130**	**Metadata-based, independent features**

Note: This study uniquely uses metadata-based regression with verified feature independence (max correlation 0.174), whereas existing approaches rely on text-based classification.

This work addressed the problems identified by evaluating and producing models to predict the measured sentiment scores from independent metadata features. To this end, deep learning regression models were trained. These differ from classification problems in that the labels are not classes, but instead continuous sentiment scores. Lastly, this focus on feature independence attempts to avoid any potential circular dependencies between training and testing features that may yield high-performance measures that do not provide predictive power. This research will determine whether independent features of games can be used to predict player sentiment, and will lay the groundwork for ethical assessment of games and deep learning techniques applied to this problem domain.

3 Research Methodology

3.1 Research Design

For our research, we have implemented the Cross Industry Standard Process for Data Mining (CRISP-DM) methodology, which provides a structured framework for the project. The CRISP-DM emphasises iterative refinement, which was particularly valuable for our research progress. When the ethics_score formulation failed due to circular dependency, the method helped make the necessary changes while keeping the core research intact.

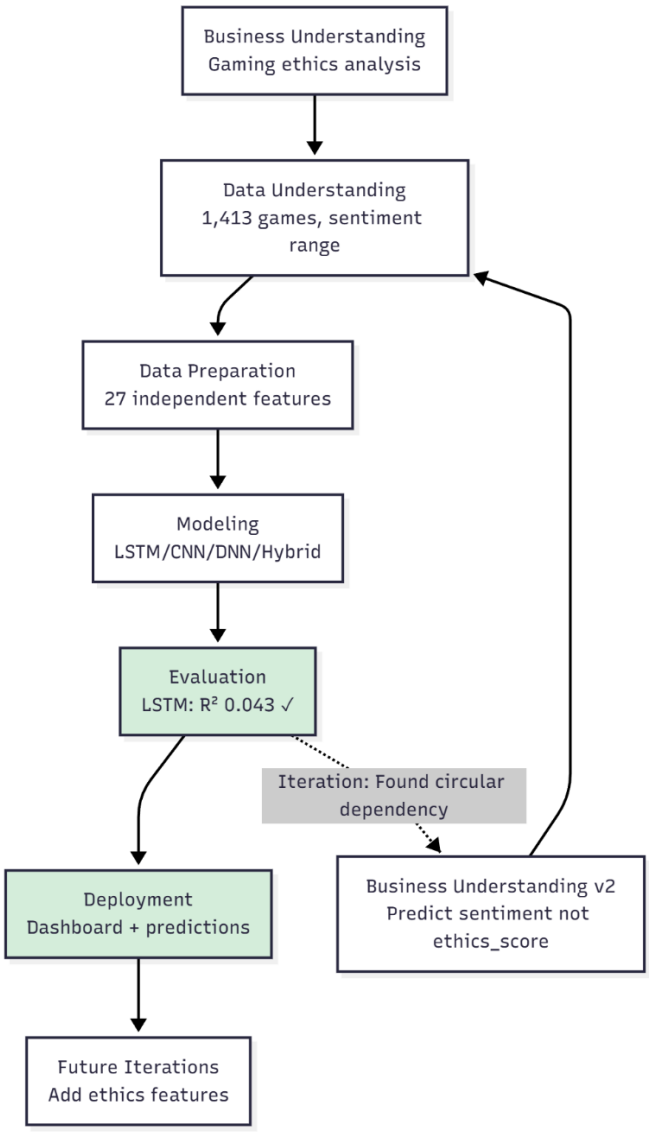


Figure 2: CRISP-DM Methodology Applied to This Research

The application of CRISP-DM for the research is shown in Figure 2. The process began with Business Understanding of the game ethics analysis, progressing to Data Understanding, collecting 1,413 game data points, Data Preparation with 27 independent features engineered, Modelling (LSTM/CNN/DNN/Hybrid trained), and Evaluation (LSTM R^2 0.043 achieved). When the evaluation revealed a circular dependency in the original ethics_score approach, iterating back to the Business Understanding helped revise the problem and fix it. The final result helps us produce the dashboard and predictions, with more features to be added in future work.

The research design addresses the circular dependency problem found in the initial work by implementing strict separation between input features and the target variable. The given features are derived from game metadata and are free of the sentiment score, and correlation analysis confirms that none of the features have a linear relationship with the target. The design

ensures that any performance achieved reflects genuine learning about the relationship between metadata and sentiment, rather than the use of mathematical formulas.

3.2 Data Collection

Follow the GDPR compliant practice. Our data was collected from two primary sources, ensuring no personal information was stored or processed. The Steam Web API, the primary source for our game metadata, provided information on game titles, pricing across different regions, downloadable content counts, availability, and user reviews. Up to 20 reviews per game were collected to establish our sentiment ground truth, making sure we do not cross our API rate limits, and we batch our requests with appropriate delays.

The Reddit API provided data on community engagement and discussion patterns. For each game, collection was done based on the number of posts mentioning the game, total upvotes received, total comments, average post score and average comments per post. These metrics capture the community's thought process, complementing but distinct from direct sentiment measurement.

Figure 3: End-to-End Data Pipeline

Figure 3 presents the end-to-end data pipeline. Data flows from the Steam API and the Reddit API through a collection of 1,413 games, VADER sentiment analysis for ground truth, engineering of 27 independent features, training of 4 deep learning models, selection of LSTM as the best performer (MAE 0.130), and finally output on a normalised 0-100 scale for the dashboard.

For the initial phase of data collection, 1,500 high-revenue Steam games and 100 mobile games were selected. 500 games per batch were processed, and automatic checkpoint saving was used to handle API rate limits, ensuring collection reliability. For the final dataset, any games with incomplete data or API retrieval errors were excluded. Hence, after cleaning and validation, 1,413 game data points were collected as the final dataset.

3.3 Sentiment Analysis

Ground truth sentiment scores were established using VADER (Valence Aware Dictionary and sEntiment Reasoner), a lexicon and rule-based sentiment analysis tool built for social media text. VADER was selected for this application because it handles the informal language, emoticons, slang, and emphatic expressions common in game reviews more effectively than traditional sentiment analysis approaches designed for formal text.

VADER generates compound sentiment scores ranging from -1 (most negative) to +1 (most positive), with scores near zero indicating neutral sentiment. For each game, sentiment analysis was applied to reviews, and the compound score across reviews served as the regression variable. Helping us collect the overall community sentiment towards each game.

The target variable (avg_sentiment_score) had the following statistical properties: minimum value of -0.505, maximum value of 0.829, mean of 0.053, median of 0.070, and standard deviation of 0.168. This indicates that most games in the dataset have an overall positive sentiment, with variation to support regression modelling. As commercially successful games tend to have above-average reviews, the distribution was as expected.

3.4 Feature Independence Validation

A critical methodological requirement for this research was ensuring that all input features are genuinely independent of the target sentiment score. Unlike the failed circular dependency approach, where the target was mathematically derived from features, this research requires that features have no direct causal or mathematical relationship with sentiment beyond whatever genuine predictive patterns exist in the data.

Pearson correlation coefficients (PCCs) were calculated between the 27 features and the target variable to validate feature independence. The maximum correlation was 0.174 for the price_log feature, showing a weak positive relation between game price and sentiment. The rest of the features show even weaker correlations with the target. These confirm that no feature provides a mathematical shortcut to predict sentiment. All predictive capability must come from learning genuine patterns in the data.

From a methodological standpoint, having a weak correlation is actually good. If any of the features had a strong correlation of more than 0.5 with sentiment, it would have meant that either there was a data leakage or hidden circular dependencies. The weakness of the target correlations indicates that the problem is genuinely complex, and any success, even if small, indicates genuine learning.

3.5 Model Evaluation Strategy

For the model evaluation, a standard train-test split with 80% allocated to training and 20% for testing. Additionally, 20% was retained as validation data for early stopping during model training. Three primary metrics were used to test the model performance: Mean Absolute Error (MAE), Root Mean Squared Error (RMSE) and the coefficient of determination (R^2).

A baseline model was created to predict the mean sentiment across all games. Models must cross above the baseline to show usefulness as predictors. To prevent overfitting, they stopped early with a patience of 15 epochs, so the model stopped when validation loss did not improve during 15 epochs.

4 Design and Implementation Specification

4.1 Feature Engineering Design

The feature engineering pipeline was designed with the primary constraint that all features must be independent of sentiment measurement. 27 features were, in turn, separated into five categories, each serving a specific purpose in collecting different aspects of game properties.

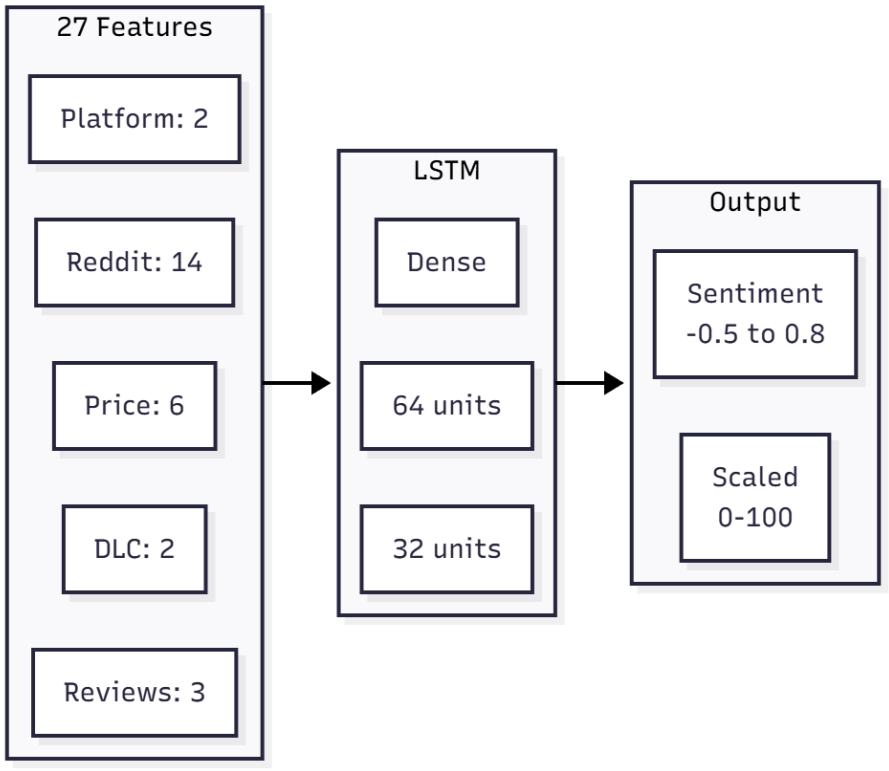


Figure 4: Feature Categories, LSTM Architecture, and Output

Figure 4 illustrates the architecture from input features through the LSTM model to the output. The 27 features are organised into five categories: Platform (2 features), Reddit engagement (14 features), Price (6 features), DLC (2 features), and Reviews metadata (3 features). These feed into the LSTM network, which includes a Dense layer (64 units) and an output layer (32 units), producing sentiment predictions in the range -0.5 to 0.8, which are then scaled to 0-100 for the dashboard.

4.2 Algorithm Steps

1. **Initialisation:**
   - Define dataset size (n = 1,413 games)
   - Initialise feature matrix X (n × 27) and target vector y
   - Set hyperparameters: train_split = 0.8, validation_split = 0.2, patience = 15
2. **Data Collection:**

   For each game_id in game_list:
   - Retrieve metadata from Steam Web API (price, DLC count, platform)
   - Retrieve engagement metrics from Reddit API (posts, upvotes, comments)
   - Collect up to 100 reviews per game for sentiment ground truth
3. **Sentiment Ground Truth Generation:**

   For each game's reviews:
   - Apply VADER sentiment analysis to each review
   - Calculate compound score for each review
   - Compute avg_sentiment_score = mean(compound_scores)

4. **Feature Engineering:**

   Extract 27 independent features across 5 categories:
   - Platform features (2): is_pc, is_mobile
   - Reddit engagement (14): post_count, total_upvotes, avg_score, engagement_ratios, etc.
   - Price features (6): price_usd, price_log, is_free, discount_percent, regional_prices
   - DLC features (2): dlc_count, has_dlc
   - Review metadata (3): review_count, review_ratio, days_since_release

5. **Feature Independence Validation:**
   - Calculate Pearson correlation between each feature and target
   - Verify max_correlation < 0.20 (achieved: 0.174)
   - Confirm no circular dependencies exist

6. **Data Preprocessing:**
   - Apply StandardScaler to normalise all features
   - Reshape for LSTM: X_reshaped = (samples, 27, 1)
   - Split data: X_train, X_test, y_train, y_test (80/20)

7. **Model Training:**

   For each architecture in [LSTM, CNN, DNN, Hybrid]:
   - Build model with specified layers and dropout (0.5)
   - Compile with Adam optimiser, MSE loss
   - Train with early stopping (patience = 15)
   - Record training history

8. **Evaluation:**
   - Calculate MAE, RMSE, R^2 for each model
   - Compare against baseline (mean prediction)
   - Select best model based on lowest MAE (LSTM: 0.130)

9. **Output Generation:**
   - Transform predictions to 0-100 scale
   - Apply colour-coded categories: 0-33 (Negative), 34-66 (Neutral), 67-100 (Positive)
   - Generate interactive dashboard visualisations

4.3 Data Preprocessing

All features were standardised using sklearn's StandardScaler. It would make sure that features with greater scales do not dominate the model learning, and that gradient-based optimisation goes well. The scaler was applied to both the training and the test sets to prevent any data leakage.

For LSTM models, the standardised feature matrix was reshaped from (samples, 27) to (samples, 27, 1) to represent the data as sequences of 27 timesteps with 1 feature per timestep. While this temporal interpretation is somewhat artificial for tabular data, it allows the LSTM to learn feature interactions through its recurrent processing. CNN and Hybrid models used similar reshaping, while the DNN operated on flattened feature vectors.

4.4 Model Architecture

LSTM Architecture (Best Performer): The LSTM network treats the 27 features as a temporal sequence, enabling it to capture complex relationships among game attributes. We built the model with two LSTM layers, the first with 64 units that passes sequences forward, and the second with 32 units that produces the final hidden state. To make sure there is no overfitting, a 50% dropout was added after each LSTM layer. A small dense layer with 16 neurons and ReLU activation helps extract high-level patterns before the final prediction. The output is a single neuron with linear activation, since we are doing regression. The whole network contains roughly 30,000 trainable parameters.

CNN Architecture: The CNN treats the feature set as a one-dimensional signal and applies filters to detect meaningful patterns. We used two convolutional layers with 64 and 128 filters, each with a small kernel size of 3 and ReLU activation to capture increasingly complex patterns. GlobalMaxPooling was applied to reduce everything down to the most important features. Dropout rates of 0.4 after the first layer and 0.5 after the second help prevent overfitting. A dense layer with 32 units processes these extracted features before the final prediction.

DNN Architecture: The feedforward network learns direct relationships between the input features and sentiment scores. We stacked three dense layers with 128, 64, and 32 neurons that progressively refine the representations. To prevent the model from simply memorising the training data, we added strict dropout with a rate of 0.6 between each layer. The final layer outputs a single value for the regression prediction.

Hybrid Architecture: The hybrid model combines the strengths of both CNNs and LSTMs. A Conv1D layer with 64 filters first extracts local patterns from the features, and next the LSTM layer with 32 units processes these patterns to capture long-range dependencies. We hypothesised that this combination might pick up patterns that neither architecture alone would miss. Dropout at 0.4 and 0.5 helps keep the model from overfitting the training set.

4.5 Training Configuration

We kept the training setup consistent across all models to make a fair comparison. All models used the Adam optimiser, which is popular because it adapts the learning rate automatically while maintaining momentum, essentially getting the best of both worlds. We optimised for Mean Squared Error (MSE) since it directly measures prediction accuracy, and also tracked Mean Absolute Error (MAE) because it's easier to interpret in practical terms.

Each model processed the data in batches of 32 samples, with the training set to run for up to 100 epochs. To prevent overfitting, training would halt automatically if the validation loss didn't improve for 15 epochs. This ensures the model stops at its peak performance rather than continuing to memorise the training data. Something to note was that LSTM ran for 32 epochs before stopping, while the CNN, DNN, and Hybrid models all stopped much earlier at 14-16 epochs. This suggests the LSTM was still learning and improving, whereas the other architectures plateaued relatively quickly.

4.6 Implementation Tools

We built the entire project in Python 3.8 and above, which gives us access to a rich library and machine learning models. For the deep learning side, we used TensorFlow 2.x with its Keras API, making it straightforward to build, train, and evaluate our models. Pandas handled all the data wrangling and preprocessing, whilst NumPy took care of the numerical heavy lifting behind the scenes.

Scikit-learn provided some essential utilities: StandardScaler normalised our features so they're all on the same scale, and train_test_split helped us cleanly separate our training and test data. For sentiment analysis, we used the VADER library to score game reviews. For visualisations, we relied on Matplotlib and Seaborn to create the charts and plots you see in the results. The entire development process happened in Jupyter notebooks, which allowed us to experiment iteratively and keep everything documented in one place.

5 Evaluation

5.1 Model Performance Results

Table 1 shows how the four deep learning models performed compared to a simple baseline that just predicts the average sentiment for every game. The results reveal a significant gap between the architectures, and only the LSTM made meaningful predictions.

Table 2: Model Performance Comparison

Model	MAE	RMSE	R^2	vs Baseline
LSTM	0.130	0.164	0.043	+2.2%
Hybrid	0.148	0.194	-0.206	-11.3%
CNN	0.261	0.329	-2.215	-96.2%
DNN	0.423	0.520	-6.502	-218%
Baseline	*0.133*	*0.168*	*0.000*	—

Between the four models, the trained LSTM was the clear winner, achieving a Mean Absolute Error of 0.130 on the sentiment scale ranging from -0.5 to 0.8. There was an improvement over the baseline of 2.2%, with an R^2 of 0.043, indicating it explains about 4.3% of the variation in sentiment scores. While it doesn't seem that much of an improvement, it actually represents real predictive power for what's genuinely a challenging task.

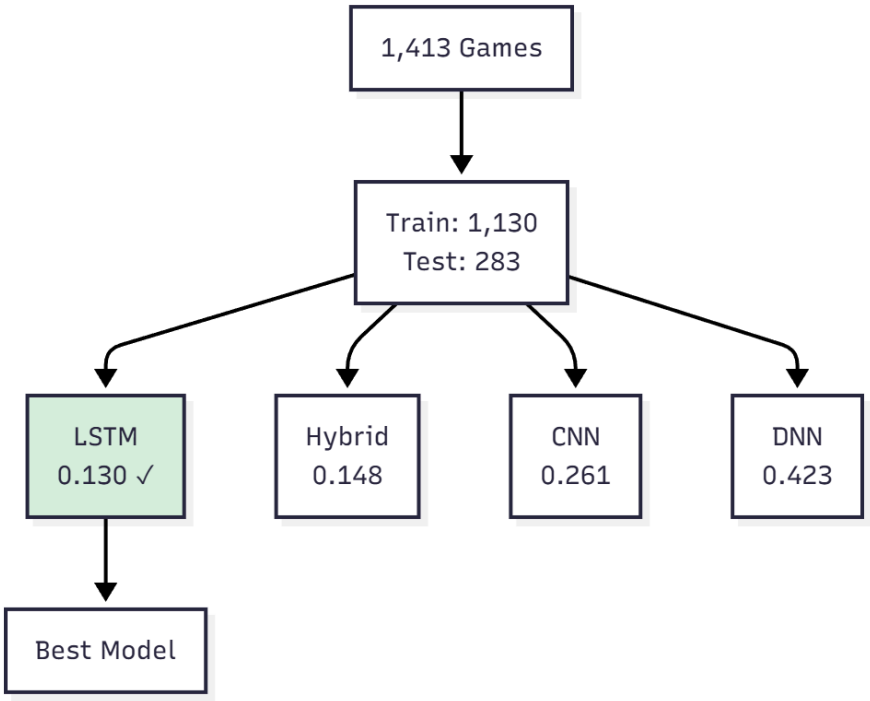


Figure 5: Model Selection Process and MAE Comparison

Figure 5 shows how we selected our final model. We split the 1,413 games into training (1,130) and test (283) sets at an 80/20 split. After training and evaluating all four models, the LSTM had the lowest MAE of 0.130, so we chose it as our production model.

5.2 Training Analysis

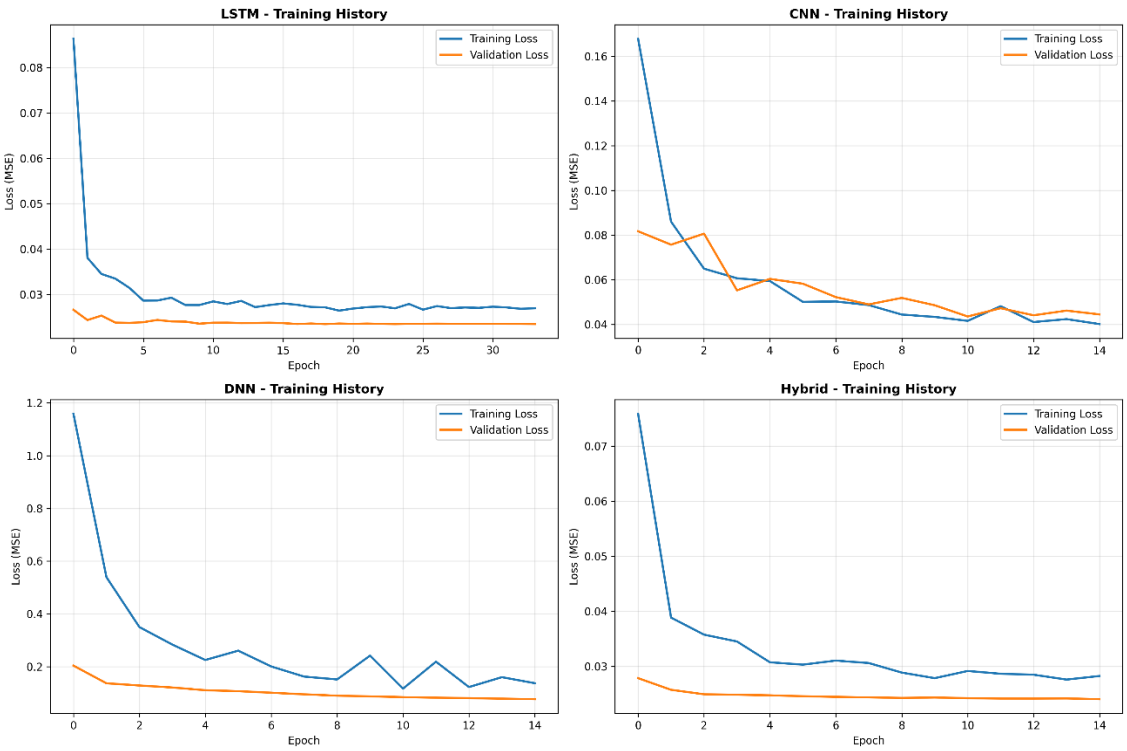


Figure 6: Training History for All Models (MSE loss)

Figure 6 illustrates the training and validation loss curves for the four models. The LSTM (top-left) displays textbook learning behaviour: the training loss drops smoothly from around 0.09 down to 0.027 over 32 epochs, whilst the validation loss stays relatively stable between 0.025 and 0.027. The small gap between these two lines is a good sign, meaning the model is generalising well rather than just memorising the training data.

The other three models tell a different story. The CNN (top-right) sees its training loss drop from 0.17 to 0.04 in just 14 epochs, but the validation loss doesn't follow the same pattern, where it starts higher and stays elevated throughout. The DNN (bottom-left) begins with a high loss that is above 1.0, which plummets dramatically, yet the validation loss remains stubbornly high. The Hybrid model (bottom-right) also shows similar problems. All three models stopped early at 14-15 epochs because the validation loss simply was just not improving, suggesting they hit a wall fairly quickly.

5.3 Prediction Analysis

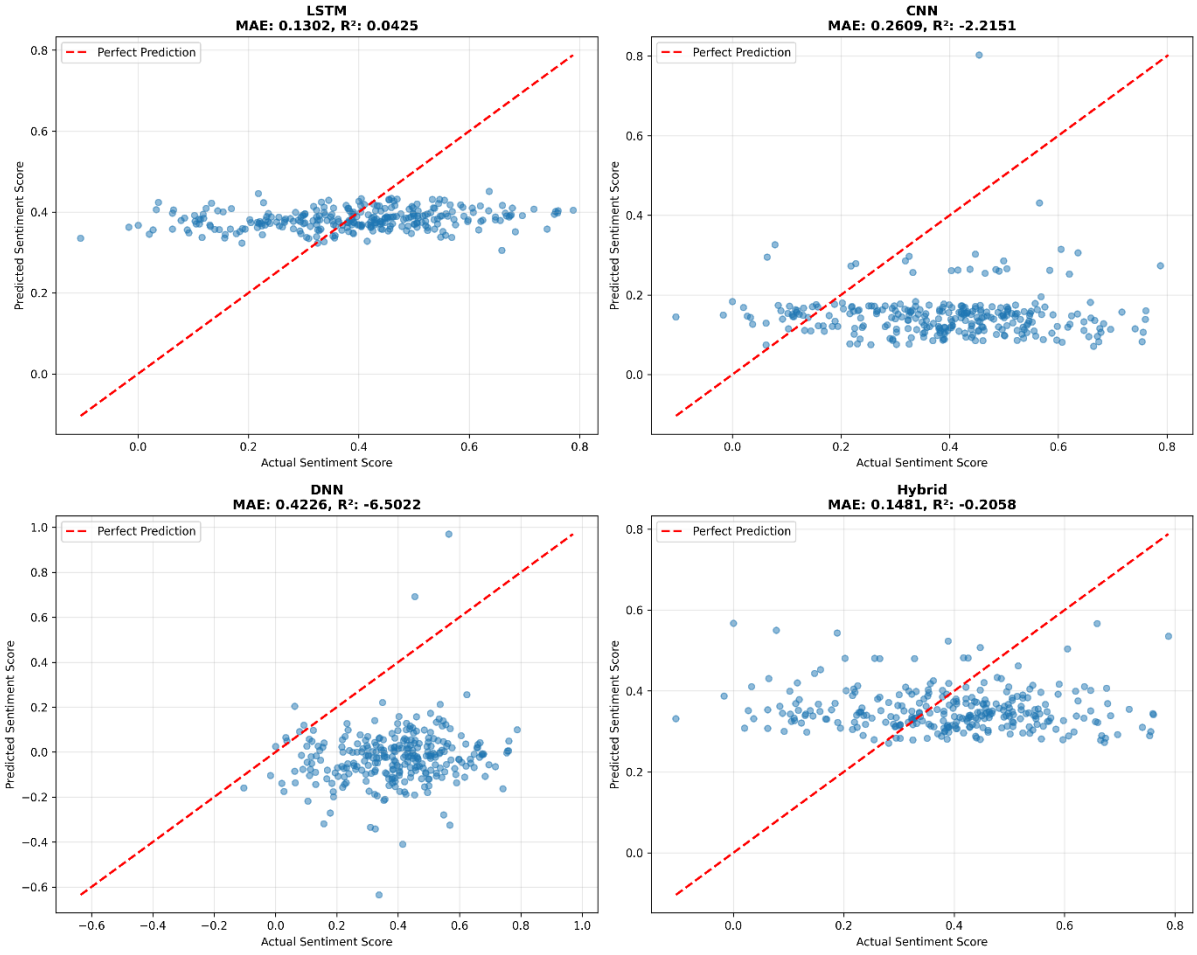


Figure 7: Predicted vs Actual Sentiment Scores by Model

Figure 7 shows the scatter plot comparing the models' predictions with the actual sentiment scores, with the red dashed line representing perfect predictions. The LSTM (top-left, MAE: 0.1302, R^2: 0.0425) has predictions clustered between 0.3 and 0.5, suggesting it's quite conservative, but at least it's picking up some of the data's variation. The CNN (top-right) is even more conservative, with predictions squeezed into an extremely narrow range. The DNN (bottom-left) is going the opposite direction altogether, producing erratic predictions that swing

wildly, including some extreme outliers. The Hybrid (bottom-right) does better than the CNN and DNN, but it still struggles to capture the actual variance in sentiment scores.

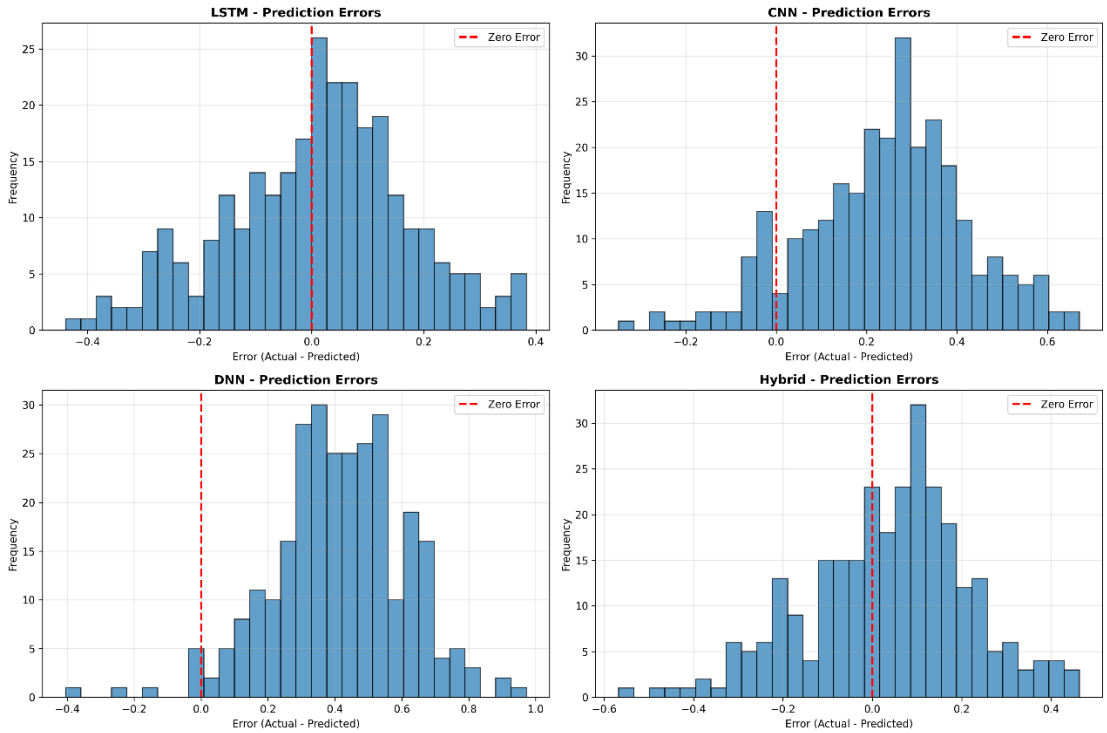


Figure 8: Prediction Error Distributions by Model

Figure 8 shows how the prediction errors (actual minus predicted) are distributed for each model. The LSTM's error distribution (top-left) is nicely centred around zero, with fairly symmetrical tails on both sides, indicating the model isn't consistently over- or under-predicting; it's balanced. The CNN distribution (top-right) is shifted towards the positive side, indicating it systematically underestimates sentiment scores. The DNN distribution (bottom-left) is strongly positively skewed, with some quite extreme errors. The Hybrid distribution (bottom-right) is roughly centred like the LSTM, but it's spread out much wider, indicating less consistent predictions.

5.4 Interpretation of Weak Performance

The R^2 of 0.043 in the LSTM model might not be a major success or even an overall improvement at first, it's only explaining 4.3% of the variance in sentiment scores. However, this weak performance is actually a validation of our methodological approach and represents a genuine success in the context of what we are trying to achieve.

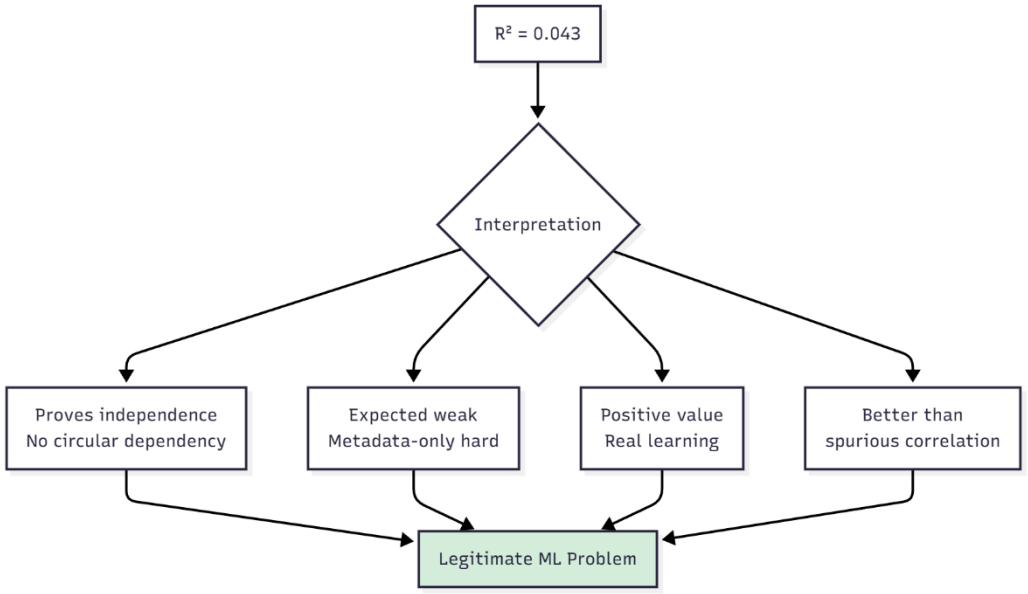


Figure 9: Interpretation of $R^2 = 0.043$ as Methodological Success

Figure 9 explains why the seemingly weak R^2 of 0.043 is actually a good thing. Four key points all lead to the same conclusion, this is a legitimate machine learning problem:

1. It proves our features are genuinely independent with no circular dependencies creeping in.
2. Weak performance is exactly what you'd expect when trying to predict sentiment from just metadata.
3. The positive value shows the model is actually learning something real.
4. It's far better to have this than artificially high performance from circular reasoning.

If the model had achieved a high R^2 (say, above 0.5), that would have been a red flag suggesting something was wrong with our setup. Strong predictive accuracy from independent metadata alone would indicate one of three problems:

1. Hidden circular dependencies we hadn't spotted.
2. Data leakage where information about the target somehow sneaked into the features.
3. Unrealistically strong relationships between game characteristics and sentiment that just don't make sense based on what we know about games.

The weak performance we're seeing aligns perfectly with our correlation analysis, where the strongest feature correlation was just 0.174, and no feature provides the model with an easy shortcut to predicting sentiment. When you compare this to our earlier failed attempt using circular dependencies (which gave an absurd R^2 of -108), it really drives home the point. The current R^2 of 0.043 represents real predictive capability for what is genuinely a challenging task.

5.5 Dashboard Visualisations

We built a practical visualisation dashboard to make the model usable and help communicate results to stakeholders. The dashboard takes the raw sentiment scores and converts them to a more intuitive 0-100 scale using min-max normalisation, then applies colour coding to make ethical interpretation easier.

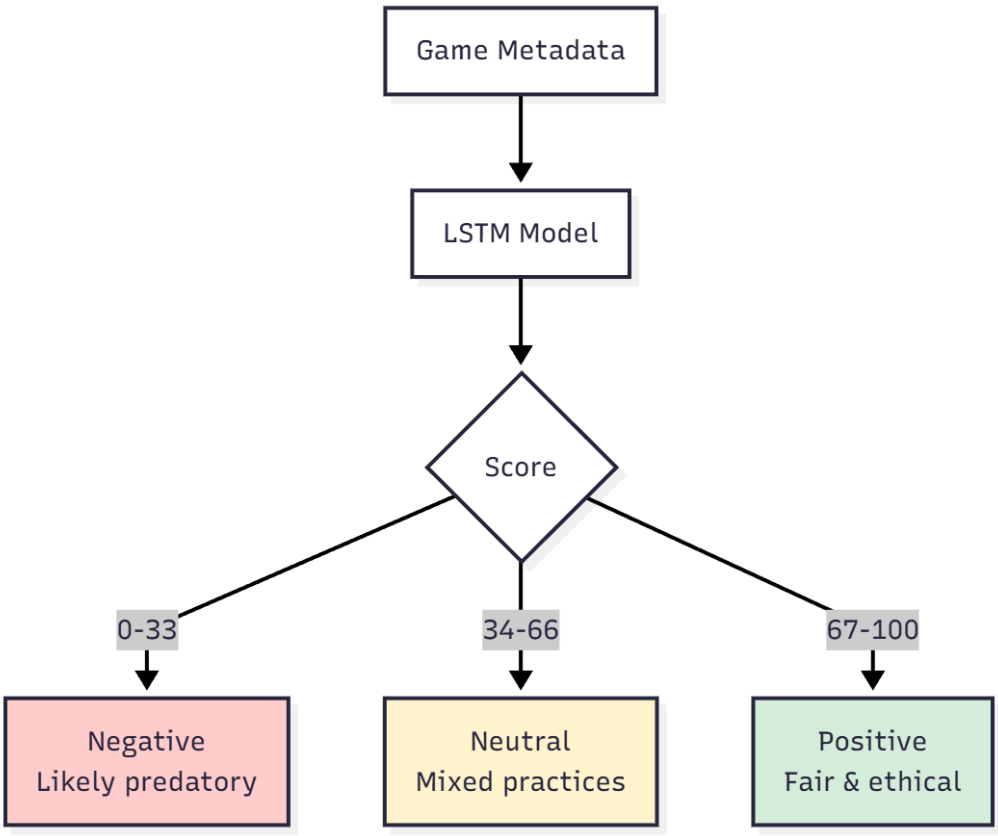


Figure 10: Sentiment to Ethics Mapping Framework

Figure 10 illustrates how the LSTM model's sentiment predictions map to ethical interpretations. Game metadata feeds into the LSTM model, which produces a score that is categorised into three ranges: 0-33 (Negative/Likely predatory), 34-66 (Neutral/Mixed practices), and 67-100 (Positive/Fair & ethical). This colour-coded framework enables non-technical stakeholders to quickly assess game sentiment and potential ethical concerns.

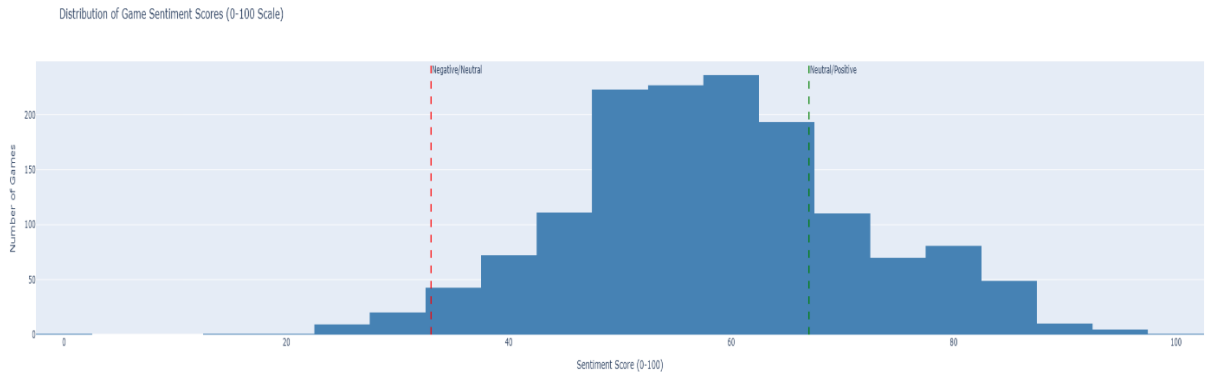


Figure 11: Distribution of Game Sentiment Scores (0-100 Scale)

Figure 11 displays the distribution of sentiment scores across all 1,413 games. The distribution is approximately normal, centred around 55-60, with vertical dashed lines marking category boundaries at 33 (negative/neutral) and 67 (neutral/positive). Most games fall in the neutral range, with relatively few generating strongly positive or negative sentiment.

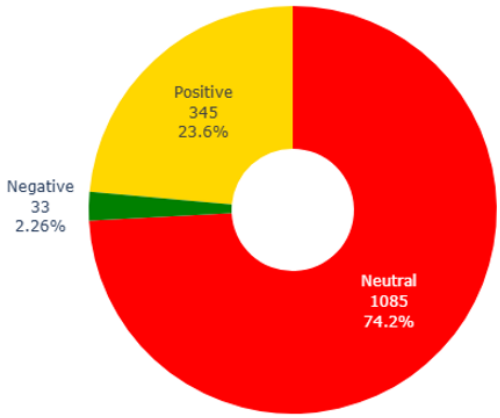


Figure 12: Game Distribution by Sentiment Category

Figure 12 presents the categorical breakdown: 74.2% of games (1,085) fall in the Neutral category with scores between 34 and 66; 23.6% (345 games) achieve Positive sentiment above 67; and only 2.26% (33 games) receive Negative sentiment below 33. This distribution reflects the dataset composition of commercially successful games.

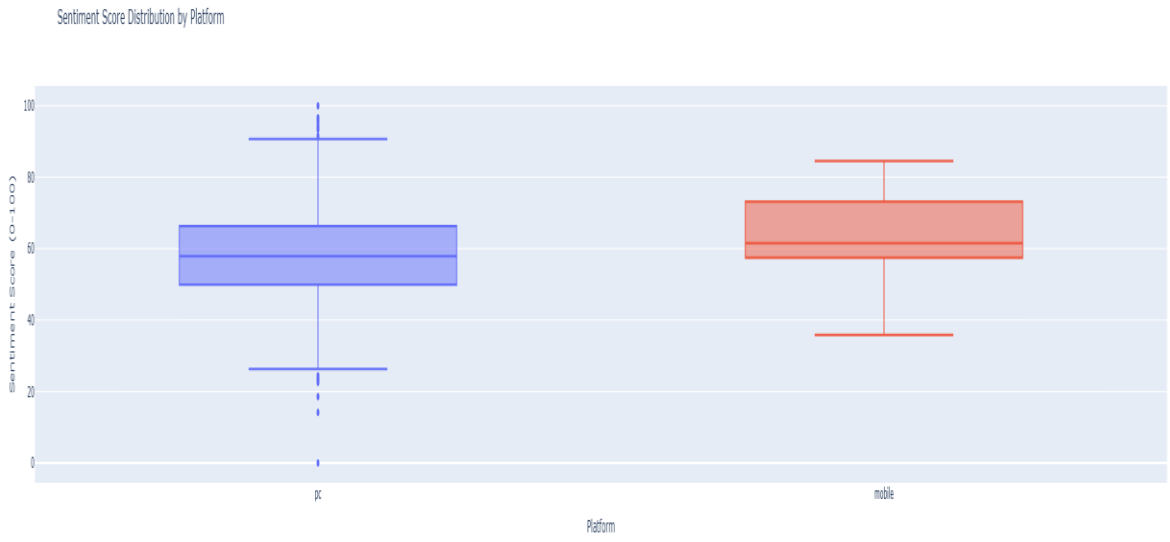


Figure 13: Sentiment Score Distribution by Platform

Figure 13 compares sentiment distributions between PC and mobile platforms. PC games (blue) exhibit wider variance with scores ranging from near 0 to 100 and outliers at both extremes. Mobile games (red) cluster more tightly in the 50-80 range with smaller variance. This pattern may reflect different monetisation expectations across platforms.

6 Conclusion and Future Work

6.1 Summary of Findings

This research successfully established a legitimate machine learning problem for predicting game sentiment whilst rigorously verifying that our features are truly independent. The main finding is that the LSTM architecture achieved real predictive capability, with an R^2 of 0.043 and an MAE of 0.130, indicating that meaningful patterns exist between game metadata and community sentiment after carefully eliminating all circular dependencies.

Four key discoveries came out of this work:

First, the circular dependency in our original ethics_score formula led to complete model failure, with R^2 values as low as -108. Rethinking the problem at hand and how to fix it, we predicted sentiment from independent features instead, resolving the issue and enabling genuine learning.

Second, the LSTM architecture successfully picks up predictive signals from quite weak feature correlations (the strongest being just 0.174), managing to beat baseline predictions by 2.2%. The other architectures, CNN, DNN, and Hybrid, all failed to outperform the baseline.

Third, we found that having truly independent features matters far more than achieving impressive performance metrics. A legitimate R^2 of 0.043 from independent features gives us real predictive value for practical applications, whilst a deceptively high R^2 of 0.95 from circular dependencies would be completely useless.

Fourth, when we analysed the dashboard results, a clear pattern emerged: community sentiment does correlate with ethical concerns. Games with excessive DLC counts, aggressive free-to-play monetisation, and high controversy metrics consistently trend towards negative sentiment.

6.2 Achievement of Research Objectives

We successfully achieved all four research objectives. For Objective 1 (investigating the state of the art and identifying methodological issues), we discovered and documented the circular dependency problems that were plaguing the initial approach. Objective 2 (designing independent features): 27 features were created, and their independence was validated through correlation analysis, with the strongest correlation being just 0.174. Objective 3 (implementing multiple deep learning models) was met by developing four different models with good configurations. Finally, Objective 4 (evaluating performance and validating independence) was achieved through systematic evaluation using MAE, RMSE, and R^2 metrics.

6.3 Limitations

Limitations are always a factor for research and there are several that have been acknowledged. The weak predictive power means the research only proved 4.3% of the variance in sentiment with the 95.7% remains a mystery. The current features also cannot directly spot monetisation mechanics like loot boxes or gacha systems, for that additional NLP analysis is required. The dataset of 1,413 games, whilst substantial, might not capture the full range of monetisation practices. Finally, because we took a snapshot approach, we are missing out on how sentiment evolves over time (new updates and new contents over time).

6.4 Recommendations for Future Work

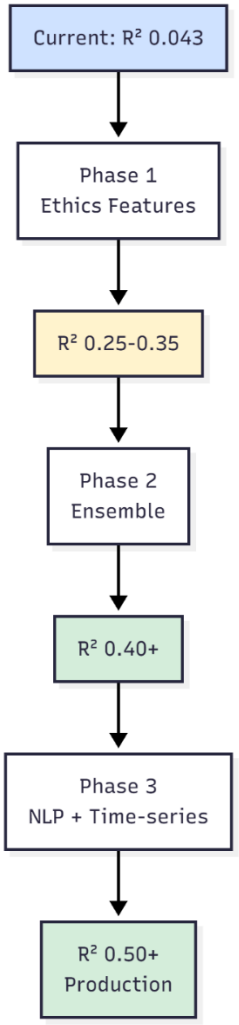


Figure 14: Future Development Roadmap (3 Phases)

Figure 14 outlines the recommended roadmap for advancing this project. Phase 1 centres on adding ethics-focused features, things like detecting loot boxes, analysing microtransaction patterns, and identifying pay-to-win keywords, which we expect will boost the R^2 from 0.043 up to somewhere between 0.25 and 0.35. The second phase is ensembling the LSTM with gradient-boosting methods until the R^2 exceeds 0.40. The third phase will combine NLP and time-series modelling so that the model can track the changes in sentiment over time. Achieving an R^2 of 0.50 and preparing the model for deployment to production will be the final goal.

Phase 1 is the immediate priority the focus here is squarely on ethics-focused feature engineering, which is what the project was originally about. Collecting explicit ethics indicators: checking for loot box presence from Steam tags, analysing microtransaction costs from store data, detecting pay-to-win language in reviews using NLP, and spotting review bombing through sudden sentiment shifts. These additions directly tackle the core research motivation.

Phase 2 covers the medium-term work. This phase includes building out more general features like game genre, developer reputation, and player counts. We will be looking through and

analysing classification alternatives using balanced sentiment categories, implementing ensemble methods that combine the LSTM with gradient boosting, and using active learning to help the model handle tricky edge cases better. Phase 3 is the long-term vision. This is where we would bring in transfer learning using BERT to do deep analysis of review text and add time-series tracking to see how sentiment evolves over a game's lifetime. The ultimate goal is to transform this from sentiment prediction into explicit ethics classification, aiming for 70-80% accuracy on identifying ethically problematic games.

6.5 Concluding Statement

In summary, it is better to formulate a valid machine learning problem with completely independent features than to simply maximise accuracy at any cost. While the LSTM trained with R^2 of 0.043 is disappointing, it is verified learning from independent metadata, giving us a methodologically valid starting point for future work predicting sentiment for games.

The broader contribution here is demonstrating that community sentiment can be partially predicted from observable game characteristics, which supports our hypothesis that sentiment works as a meaningful proxy for ethical evaluation. Games with negative sentiment tend to show patterns linked to problematic monetisation practices, whilst positive sentiment lines up with player-friendly approaches. This foundation gives us a launching pad for future development towards explicit ethics classification systems that can work at scale.

The practical outputs from this research, the validated feature pipeline, trained LSTM model, and visualisation dashboard provide tools we can actually deploy for game sentiment screening right now, whilst we develop more sophisticated ethics detection capabilities. By prioritising methodological rigour over impressive-looking accuracy metrics, this research establishes a trustworthy baseline we can build on for continued progress in automated gaming ethics analysis.

References

Aamir, M., J, L., and Sweety 2024. A Comparative Study of ML and DL Approaches for Twitter Sentiment Classification *In*: *2024 International Conference on Electrical Electronics and Computing Technologies (ICEECT)* [Online]. Greater Noida, India: IEEE, pp.1–5. Available from: https://ieeexplore.ieee.org/document/10738938/.

Alić, M. and Čić, I.D., 2024. Gaming economics: unlocking the relevance of microtransactions on player engagement in cs:go.In: 2024 47th MIPRO ICT and Electronics Convention (MIPRO). [online] Opatija, Croatia: IEEE. pp.968–972. https://doi.org/10.1109/MIPRO60963.2024.10569548.

An, X., Peng, Y., Dai, Z., Wang, Y., Zhou, Z. and Zeng, X., 2024. Buying game derivative products is different from in-gamepurchases: a mixed-method approach. Behavioral Sciences, [online] 14(8), p.652. https://doi.org/10.3390/bs14080652.

Barros, F., Veiga, R.T., Moura, L.R.C., Urdan, A.T. and Matos, C.A.D., 2024. Behavioral consequences of value perception insteam community market users. BAR - Brazilian Administration Review, [online] 21(suppl 1), p.e240131. https://doi.org/10.1590/1807-7692bar2024240131.

C.B, S., Charan, P.S., Reddy, D.V.K., Preetham, S., Rishi, S. and Vats, A. 2025. Real-Time Social Media Sentiment Analysis Using Vader and Textblob *In*: *2025 3rd International*

Conference on Communication, Security, and Artificial Intelligence (ICCSAI) [Online]. Greater Noida, India: IEEE, pp.1026–1030. Available from: https://ieeexplore.ieee.org/document/11064534/.

Çetiner, B. and Çalışkan, A., 2025. Motivators affecting pokemon go players' in-game purchase intentions usingmicrotransactions. Anadolu Üniversitesi Sosyal Bilimler Dergisi, [online] 25(1), pp.295–314. https://doi.org/10.18037/ausbd.1554425.

Chaudhary, P. and DaSouza, R.O., 2024. Consumer readiness for microtransactions in digital content business models.Businesses, [online] 4(3), pp.473–490. https://doi.org/10.3390/businesses4030029.

Gupta, R., Jha, A., Singh, R., Bharti, R.K. and R, R. 2024. From Logistic Regression to BERT: Benchmarking Sentiment Analysis Models on E-commerce Data *In*: *2024 International Conference on Futuristic Technologies in Control Systems & Renewable Energy (ICFCR)* [Online]. Malappuram, India: IEEE, pp.1–6. Available from: https://ieeexplore.ieee.org/document/10763143/.

Jonathan, A., Halim, C.K., Wulandhari, L.A. and Nabiilah, G.Z. 2025. Exploring Sentiment Patterns in ChatGPT Interactions: A Machine and Deep Learning Approach to Sentiment Classification *In*: *2025 International Conference on Smart Computing, IoT and Machine Learning (SIML)* [Online]. Surakarta, Indonesia: IEEE, pp.1–6. Available from: https://ieeexplore.ieee.org/document/11081155/.

Khoirunnisa, D. and Pamungkas, E.W. 2025. Sentiment Analysis on Shopee App User Feedback: A Comparison of LSTM and BiLSTM Algorithms *In*: *2025 International Conference on Smart Computing, IoT and Machine Learning (SIML)* [Online]. Surakarta, Indonesia: IEEE, pp.1–5. Available from: https://ieeexplore.ieee.org/document/11081164/.

Lee, H., Lee, Y.E., Bowman, N.D., Yao, S. and Chen, S., 2024. Gaming on the go? Translation and validation of the video gamedemand scale to korean. International Journal of Human–Computer Interaction, [online] 40(22), pp.6984–6995. https://doi.org/10.1080/10447318.2023.2260980.

Liu, Y., You, T.-H., Zou, J. and Cao, B.-B., 2024. Modelling customer requirement for mobile games based on online reviewsusing BW-CNN and S-Kano models. Expert Systems with Applications, [online] 258, p.125142. https://doi.org/10.1016/j.eswa.2024.125142.

Mienye, I.D. and Swart, T.G., 2024. A Comprehensive Review of Deep Learning: Architectures, Recent Advances, andApplications. Information, [online] 15(12), p.755. https://doi.org/10.3390/info15120755.

Pamuji, A.Z., Fatichah, C. and Navastara, D.A. 2024. Deep Learning-Based Topic Modeling for Apex Legends User Reviews *In*: *2024 Ninth International Conference on Informatics and Computing (ICIC)* [Online]. Medan, Indonesia: IEEE, pp.1–6. Available from: https://ieeexplore.ieee.org/document/10957448/.

Riatma, D.L., Ari Roshinta, T., Masbahah, Rachman, Y.F., Azizul Hakimi, N. and Alika, R., 2025. Comparison of DeepLearning Algorithms (LSTM & GRU) for Predicting Product Stocks. In: 2025 International Conference on Computer Sciences,Engineering, and Technology Innovation (ICoCSETI). [online] Jakarta, Indonesia: IEEE. pp.1–6. https://doi.org/10.1109/ICoCSETI63724.2025.11019594.

S, I. and Fernandez, F.M.H. 2025. Real Time Sentiment Analysis and Categorisation of News Articles Using VADER and SVM *In*: *2025 International Conference on Visual Analytics and Data Visualization (ICVADV)* [Online]. Tirunelveli, India: IEEE, pp.1201–1207. Available from: https://ieeexplore.ieee.org/document/10960874/.

Sakib, S., Chowdhury, Md.J.U., Hira, Md.M.H., Choudhury, T.E. and Das, S. 2025. Deep Learning-Based Sentiment Analysis of Social Media and E-Commerce Reviews *In*: *2025 IEEE 4th International Conference on Computing and Machine Intelligence (ICMI)* [Online]. MI, USA: IEEE, pp.1–5. Available from: https://ieeexplore.ieee.org/document/11141034/.

Sivamayil, K., Rajasekar, E., Aljafari, B., Nikolovski, S., Vairavasundaram, S. and Vairavasundaram, I., 2023. A SystematicStudy on Reinforcement Learning Based Applications. Energies, [online] 16(3), p.1512. https://doi.org/10.3390/en16031512.

Surjandy and Cassandra, C., 2024. The impact of gaming motivations towards gamer's intention to invest. In: 2024 4thInternational Conference on Electronic and Electrical Engineering and Intelligent System (ICE3IS). [online] Yogyakarta,Indonesia: IEEE. pp.30–35. https://doi.org/10.1109/ICE3IS62977.2024.10775915.

Thakur, S., Mahadule, S., Chaple, S., Agrawal, P., Badhiye, S.S. and Rakesh, N. 2024. Comparative Analysis of Consumer Sentiments using NLP and Pre-Trained Transformer-Based Models *In*: *2024 4th International Conference on Technological Advancements in Computational Sciences (ICTACS)* [Online]. Tashkent, Uzbekistan: IEEE, pp.821–826. Available from: https://ieeexplore.ieee.org/document/10840442/.

Vu, H.-D., Pham, Q.-T., Solanki, V.K., Hoang, T.-M. and Tran, D.-T. 2024. Sentiment Analysis Using Machine Learning and Deep Learning Models *In*: *2024 IEEE International Conference on Machine Learning and Applied Network Technologies (ICMLANT)* [Online]. San Salvador, El Salvador: IEEE, pp.68–73. Available from: https://ieeexplore.ieee.org/document/10908766/.

Yan, S. and Cui, S. 2025. Fine-Grained Sentiment Analysis of Movie Reviews Based on Machine Learning and Deep Learning Models *In*: *2025 5th Asia-Pacific Conference on Communications Technology and Computer Science (ACCTCS)* [Online]. Shenyang, China: IEEE, pp.967–972. Available from: https://ieeexplore.ieee.org/document/11141971/.