

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: 888.....

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 Supplements 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

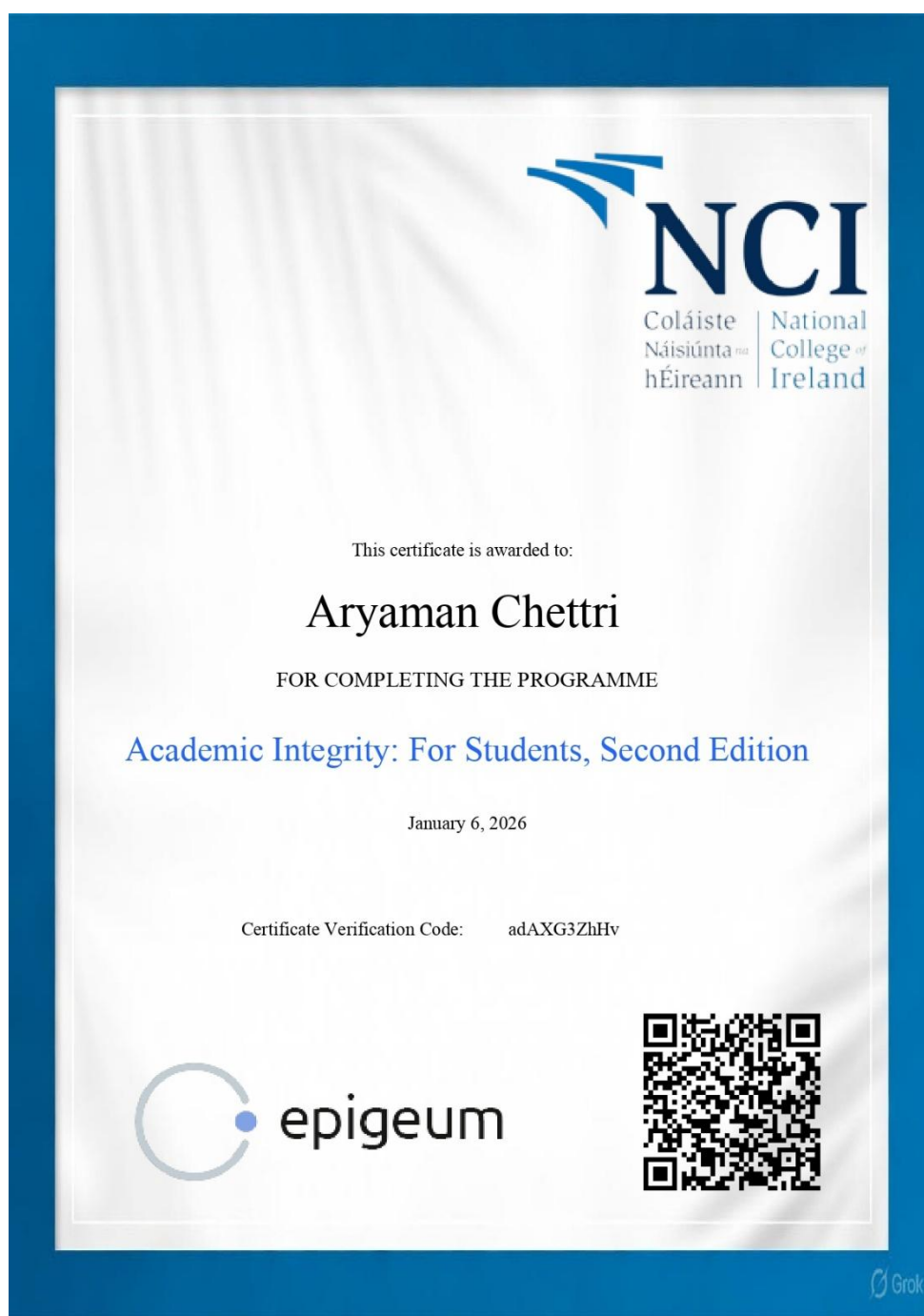
Description of AI Usage

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 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

Configuration Manual

MSc Research Project
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: ...888..... **Page Count:**...11.....

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):	

Configuration Manual

Aryaman Chettri
Student ID: 22135511

This configuration manual provides detailed information to ensure reproducibility and consistency across environments. It contains all the specific required software tools, settings, and configurations required to replicate the experimental setup for training and evaluating the proposed deep learning sentiment prediction framework.

1. System Requirements

This project was developed and implemented in multiple Jupyter Notebooks using Visual Studio Code. The following hardware and software requirements are necessary for replication.

1.1 Hardware Requirements

CPU:

- Minimum: Dual-core 2.0 GHz
- Recommended: Quad-core 3.0 GHz or higher

GPU:

- Minimum: None required (CPU-only training will work)
- Recommended: NVIDIA GPU with CUDA support (provides 10x faster training)

RAM:

- Minimum: 8 GB
- Recommended: 16 GB or more

Storage:

- Minimum: 5 GB free space
- Recommended: 10 GB or more (SSD preferred)

1.2 Software Requirements

Operating System:

- Windows 10/11 (64-bit)
- macOS (Ventura or later)
- Linux (Ubuntu 20.04 or later)

Python Environment:

- Python Version: Python 3.8 or higher (3.10 recommended)
- Environment Managers: venv (built-in) or Anaconda/Miniconda

Required Python Libraries:

Purpose	Libraries
Data handling & numerical computations	pandas>=1.3.0, numpy>=1.21.0
Deep learning & model training	tensorflow>=2.8.0, scikit-learn>=1.0.0
Sentiment analysis	vaderSentiment>=3.3.2
Plotting & visualisation	matplotlib>=3.4.0, seaborn>=0.11.0, plotly>=5.0.0
API access	requests>=2.26.0, praw>=7.5.0
Jupyter environment	jupyter>=1.0.0, ipykernel>=6.4.0

Table 1. Required Libraries

2. Code Editor Setup

Use Visual Studio Code or any other code editor like PyCharm. Download Visual Studio Code from: <https://code.visualstudio.com/>.

3. Data Gathering

This section guides you on how to collect and prepare the dataset from initial collection to final processing stage.

3.1 Dataset Sources

The data used in this project was collected from two primary sources:

Steam Web API:

Provides game metadata including pricing, DLC counts, and user reviews. Up to 20 reviews per game were collected for VADER sentiment analysis to establish ground truth sentiment scores.

Reddit API:

Provides community engagement metrics including posts, upvotes, and comments for each game.

3.2 API Configuration

Reddit API Setup:

To collect Reddit data, create a Reddit application at <https://www.reddit.com/prefs/apps>. Select 'script' for personal use. Figure 2 shows the Reddit application creation page.

create application

By creating an app, you agree to [Reddit's Developer Terms](#) and [Data Api Terms](#). You must also register to use the API.

name

☒ web app A web based application
☐ installed app An app intended for installation, such as on a mobile phone
☐ script Script for personal use. Will only have access to the developers accounts

description

about url

redirect uri

☐ I'm not a robot 
reCAPTCHA
[Privacy](#) - [Terms](#)

Figure 2. Reddit API Application Creation Page

Steam API Setup:

Register for a Steam Web API key at <https://steamcommunity.com/dev/apikey>. Figure 3 shows the Steam API key registration page.

Register Steam Web API Key

Your Steam Web API Key

Key: [REDACTED]

Domain Name: [REDACTED]

Figure 3. Steam Web API Key Registration Page

3.3 Data Collection Pipeline

The data collection pipeline is initialized with directory structure creation and VADER sentiment analysis setup. Figure 4 shows the pipeline initialization output.

```
Base directory: c:\Github\Ethics_analysis
Created/verified directory: c:\Github\Ethics_analysis\data\raw
Created/verified directory: c:\Github\Ethics_analysis\data\processed
Created/verified directory: c:\Github\Ethics_analysis\data\checkpoints
Created/verified directory: c:\Github\Ethics_analysis\data\features
[OK] Robust Gaming Ethics Data Collection Pipeline Initialized
Quality filters enabled: Price validation, Review requirements, API protection
[OK] Sentiment Analysis initialized (VADER)
```

Figure 4. Data Collection Pipeline Initialisation

Figure 5 shows the configuration settings used for the data collection process.

```

Configuration:
  Checkpoint interval: Every 50 games
  API timeout: 10s
  Steam rate limit: 1.5s delay
  Retry attempts: 3
  Require reviews: True
  Max valid price: $1000

```

Figure 5. Pipeline Configuration Settings

The collection pipeline runs in batches to avoid API rate limits. Figure 6 shows the mobile games collection process.

```

Collection Pipeline

Run data collection in batches to avoid API rate limits.

# Collect all 100 mobile games
results_mobile = collect_from_mobile_list(num_games=100, start_offset=0)

2025-11-28 14:11:54,137 - INFO - Starting Mobile collection: 100 games (offset: 0)

Collecting Mobile games 0 to 100 (100 games)
Estimated time: 8-13 minutes

Could not render content for 'application/vnd.jupyter.widget-view+json'
{"model_id":"a9293b3d875b4877b81aed13146f727c","version_major":2,"version_minor":0}

2025-11-28 14:11:59,596 - INFO - Collected 4 Reddit comments from post 1mxdy6
2025-11-28 14:12:04,089 - INFO - Collected 4 Reddit comments from post 1i3qw9h
2025-11-28 14:12:04,089 - INFO - Collected 4 Reddit comments from post 1i3qw9h
2025-11-28 14:12:04,982 - INFO - Collected 2 Reddit comments from post mz177b
2025-11-28 14:12:04,982 - INFO - Collected 2 Reddit comments from post mz177b
2025-11-28 14:12:05,769 - INFO - Collected 6 Reddit comments from post 1d4ce1m
2025-11-28 14:12:05,769 - INFO - Collected 6 Reddit comments from post 1d4ce1m
2025-11-28 14:12:10,410 - INFO - Collected 4 Reddit comments from post 1evss9l
2025-11-28 14:12:10,410 - INFO - Collected 4 Reddit comments from post 1evss9l
2025-11-28 14:12:11,182 - INFO - Collected 2 Reddit comments from post 1i83itk
2025-11-28 14:12:11,182 - INFO - Collected 2 Reddit comments from post 1i83itk
2025-11-28 14:12:11,954 - INFO - Collected 4 Reddit comments from post 1izbtzl
2025-11-28 14:12:11,954 - INFO - Collected 4 Reddit comments from post 1izbtzl
2025-11-28 14:12:16,597 - INFO - Collected 2 Reddit comments from post aik7p8
2025-11-28 14:12:16,597 - INFO - Collected 2 Reddit comments from post aik7p8
2025-11-28 14:12:17,424 - INFO - Collected 2 Reddit comments from post dmk7fi
2025-11-28 14:12:17,424 - INFO - Collected 2 Reddit comments from post dmk7fi
2025-11-28 14:12:18,164 - INFO - Collected 0 Reddit comments from post 1kqkhpz
2025-11-28 14:12:18,164 - INFO - Collected 0 Reddit comments from post 1kqkhpz
2025-11-28 14:12:22,721 - INFO - Collected 8 Reddit comments from post 1ntolim

```

Figure 6. Mobile Games Collection Pipeline

Figure 7 shows the Steam games collection in batches of 500 games with checkpoint saving.

```
# Collect Steam games in batches
# Batch 1: Games 0-500
# results_stream_1 = collect_from_steam_list(num_games=500, start_offset=0)

# Batch 2: Games 500-1000
# results_stream_2 = collect_from_steam_list(num_games=500, start_offset=500)

# Batch 3: Games 1000-1500
results_stream_3 = collect_from_steam_list(num_games=500, start_offset=1000)

[12]

... 2025-11-28 22:49:37,527 - INFO - Starting Steam collection: 500 games (offset: 1000)

Collecting Steam games 1000 to 1500 (500 games)
Checkpoints every 50 games
Estimated time: 100-125 minutes

... Could not render content for 'application/vnd.jupyter.widget-view+json'
{"model_id":"dbe0afeb44894072b461cae57cec0b42","version_major":2,"version_minor":0}

... 2025-11-28 22:49:37,978 - INFO - Found Steam app ID 274170 for 'Hotline Miami 2: Wrong Number'
2025-11-28 22:49:41,711 - INFO - Collected 15 Steam reviews for app 274170
2025-11-28 22:49:41,712 - INFO - Collected 15 Steam reviews for Hotline Miami 2: Wrong Number
2025-11-28 22:49:41,711 - INFO - Collected 15 Steam reviews for app 274170
2025-11-28 22:49:41,712 - INFO - Collected 15 Steam reviews for Hotline Miami 2: Wrong Number
2025-11-28 22:49:46,343 - INFO - Collected 3 Reddit comments from post 22im5k
2025-11-28 22:49:46,343 - INFO - Collected 3 Reddit comments from post 22im5k
2025-11-28 22:49:47,088 - INFO - Collected 3 Reddit comments from post 2yk4bk
2025-11-28 22:49:47,088 - INFO - Collected 3 Reddit comments from post 2yk4bk
2025-11-28 22:49:47,845 - INFO - Collected 3 Reddit comments from post 22imnr
2025-11-28 22:49:47,845 - INFO - Collected 3 Reddit comments from post 22imnr
2025-11-28 22:49:48,346 - INFO - Collected 9 Reddit comments for Hotline Miami 2: Wrong Number
2025-11-28 22:49:48,354 - INFO - Sentiment analysis complete for Hotline Miami 2: Wrong Number: Score=0.25, Ethics concerns=0.0%
2025-11-28 22:49:48,355 - INFO - Ethics score for Hotline Miami 2: Wrong Number: 62.4/100 (fair)
2025-11-28 22:49:48,346 - INFO - Collected 9 Reddit comments for Hotline Miami 2: Wrong Number
2025-11-28 22:49:48,354 - INFO - Sentiment analysis complete for Hotline Miami 2: Wrong Number: Score=0.25, Ethics concerns=0.0%
2025-11-28 22:49:48,355 - INFO - Ethics score for Hotline Miami 2: Wrong Number: 62.4/100 (fair)
2025-11-28 22:49:48,743 - INFO - Found Steam app ID 1547000 for 'Grand Theft Auto: San Andreas - The Definitive Edition'
2025-11-28 22:49:48,743 - INFO - Found Steam app ID 1547000 for 'Grand Theft Auto: San Andreas - The Definitive Edition'
2025-11-28 22:49:52,257 - INFO - Collected 15 Steam reviews for app 1547000
2025-11-28 22:49:52,258 - INFO - Collected 15 Steam reviews for Grand Theft Auto: San Andreas - The Definitive Edition
2025-11-28 22:49:52,257 - INFO - Collected 15 Steam reviews for app 1547000
2025-11-28 22:49:52,258 - INFO - Collected 15 Steam reviews for Grand Theft Auto: San Andreas - The Definitive Edition
2025-11-28 22:49:56,955 - INFO - Collected 2 Reddit comments from post grplx1
```

Figure 7. Steam Games Collection Pipeline (Batch Processing)

4. Execution

4.1 Project Structure

The project is organised into multiple Jupyter notebooks. Figure 8 shows the project structure in Visual Studio Code.

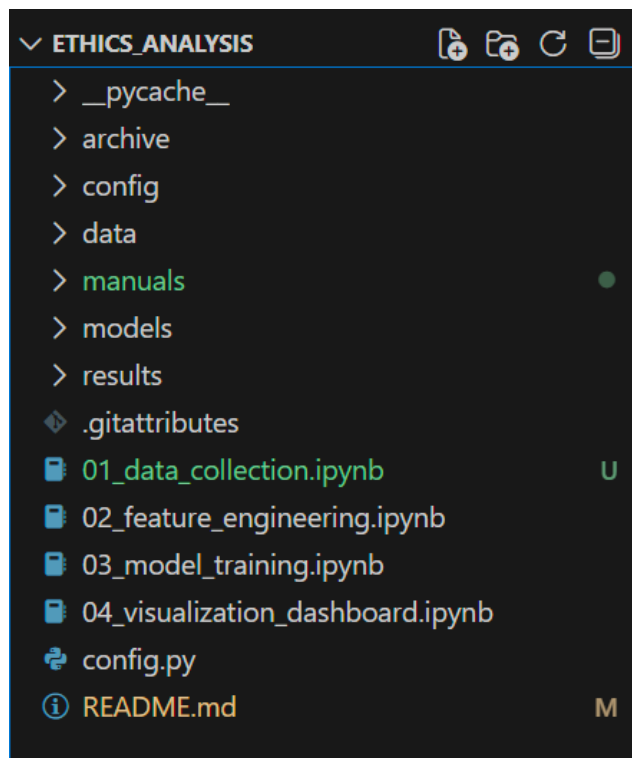


Figure 8. Project Structure in Visual Studio Code

4.2 Installation Steps

Step 1: Clone Repository

```
git clone https://github.com/Aryaman33/Ethics_analysis.git
cd Ethics_analysis
```

Step 2: Install Dependencies

```
pip install -r requirements.txt
```

4.3 Notebook Execution Sequence

Notebook	Description
01_data_collection.ipynb	Steam + Reddit data collection (~2-4 hours)
02_feature_engineering.ipynb	Feature engineering - 27 independent features (~5 min)
03_model_training.ipynb	Train LSTM, CNN, DNN, Hybrid models (~15-30 min)
04_visualization_dashboard.ipynb	Generate 0-100 color-coded dashboard (~2 min)

Table 2. Notebook Execution Sequence

5. Feature Engineering

The feature engineering notebook creates 27 independent features from the collected data. Figure 9 shows the complete list of features.

```
Using 27 features

Feature list:
1. reddit_posts
2. reddit_upvotes
3. reddit_comments
4. posts_log
5. upvotes_log
6. comments_log
7. upvotes_per_post
8. comments_per_post
9. comments_per_upvote
10. high_engagement
11. low_engagement
12. platform_pc
13. platform_mobile
14. engagement_intensity
15. community_activity
16. controversy_index
17. price_log
18. is_free
19. price_tier
20. dlc_count
21. has_dlc
22. heavy_dlc
23. review_score
24. has_reviews
25. review_category
26. total_reviews_analyzed
27. reddit_comments_collected
```

Figure 9. Complete List of 27 Independent Features

Figure 10 shows the modelling dataset summary with sentiment score distribution.

```

Modeling Dataset Summary:
  Samples: 1413
  Features: 27
  Target: avg_sentiment_score (0-1 continuous)

Sentiment Score Distribution:
  Min: -0.505
  Max: 0.829
  Mean: 0.385
  Median: 0.401
  Std Dev: 0.163
  Variance: 0.0267
  Good variance for prediction

```

Figure 10. Modeling Dataset Summary (1,413 samples, 27 features)

Figure 11 shows the top 10 features correlated with sentiment, confirming feature independence (max correlation: 0.174).

```

Top 10 Features Correlated with Sentiment:
=====
Feature                                Correlation
-----
price_log                             0.1570
total_reviews_analyzed                0.1482
price_tier                            0.1302
low_engagement                        0.0945
controversy_index                     0.0629
has_dlc                              0.0259
comments_per_upvote                   0.0231
reddit_comments_collected            -0.0027
posts_log                             -0.0383
dlc_count                             -0.0404

Max correlation: 0.1739
Weak correlations - moderate prediction difficulty

```

Figure 11. Top 10 Features Correlated with Sentiment (Max: 0.174)

6. Model Training

Four deep learning regression models were trained: LSTM, CNN, DNN, and Hybrid. Figure 12 shows the model performance summary.

```

Model Performance Summary:
=====
Model          MAE      RMSE      R2      vs Baseline
-----
LSTM           0.1302   0.1643   0.0425    2.0%
CNN            0.2609   0.3010  -2.2151   -96.3%
DNN            0.4226   0.4598  -6.5022   -217.9%
Hybrid         0.1481   0.1843  -0.2058   -11.4%
=====

Best Model: LSTM
MAE: 0.1302
R2: 0.0425

```

Figure 12. Model Performance Summary - LSTM Best Model (MAE: 0.1302, R²: 0.0425)

Table 3 provides detailed model performance metrics:

Model	MAE	RMSE	R ²	vs Baseline
LSTM	0.1302	0.1643	0.0425	+2.0%
Hybrid	0.1481	0.1843	-0.2058	-11.4%
CNN	0.2609	0.3010	-2.2151	-96.3%
DNN	0.4226	0.4598	-6.5022	-217.9%

Table 3. Model Performance Comparison

7. Model Predictions

Figure 13 shows sample predictions from the LSTM model on the test set.

```

Sample Predictions from LSTM (Best Model):
=====
Game Name          Actual   Predicted   Error
-----
Enshrouded         0.756    0.401     0.355
Call of War        0.178    0.346     0.168
Pathfinder: Wrath of the Righteous ...  0.403    0.417     0.014
Spirit City: Lofi Sessions  0.761    0.403     0.358
CloverPit          0.572    0.382     0.190
Torchlight II     0.328    0.391     0.063
Forza Horizon 4    0.357    0.342     0.015
Battle Brothers    0.325    0.356     0.031
No I'm not a Human 0.216    0.374     0.158
Screen Cat         0.369    0.355     0.014
Cyberpunk 2077     0.356    0.336     0.020
Caliber            0.337    0.327     0.010
EARTH DEFENSE FORCE 6 0.119    0.385     0.266
Ultimate Admiral: Dreadnoughts  0.092    0.349     0.257
War Selection       0.131    0.365     0.234
SILENT HILL 2      0.167    0.358     0.191
Tom Clancy's Ghost Recon® Wildlands  0.404    0.365     0.039
Spacewar           0.542    0.371     0.171
Barony             0.295    0.387     0.092
Abiotic Factor     0.696    0.391     0.305
=====

```

Figure 13. Sample Predictions from LSTM (Best Model)

Figure 14 shows the best predictions (lowest error) and worst predictions (highest error).

Best Predictions (Lowest Error):			
Game Name	Actual	Predicted	Error
Watch_Dogs	0.325	0.323	0.002
Marvel's Midnight Suns	0.382	0.380	0.002
Surviving Mars	0.403	0.400	0.003
Age of Empires: Definitive Edition	0.370	0.367	0.003
Ryse: Son of Rome	0.390	0.395	0.005
Insurgency: Sandstorm	0.386	0.381	0.005
DARKEST DAYS	0.379	0.374	0.005
Quake Live	0.388	0.382	0.006
The Riftbreaker	0.426	0.420	0.006
SAO Utils 2: Progressive	0.396	0.389	0.007
Worst Predictions (Highest Error):			
Game Name	Actual	Predicted	Error
Red Dead Online	-0.104	0.336	0.440
Holdfast: Nations At War	0.037	0.424	0.387
Dorfromantik	0.741	0.358	0.383
Maktala	0.788	0.405	0.383
Resident Evil 2	-0.017	0.363	0.380
World of Tanks Blitz	0.033	0.406	0.373
DEAD OR ALIVE Xtreme Venus Vacation	0.000	0.367	0.367
Tainted Grail: The Fall of Avalon	0.760	0.398	0.362
BALL x PIT	0.754	0.394	0.360

Figure 14. Best Predictions (Lowest Error) and Worst Predictions (Highest Error)

8. Visualisation Dashboard

The visualisation dashboard generates colour-coded sentiment scores on a 0-100 scale. Figure 15 shows the scatter plot visualisation of all 1,463 games.

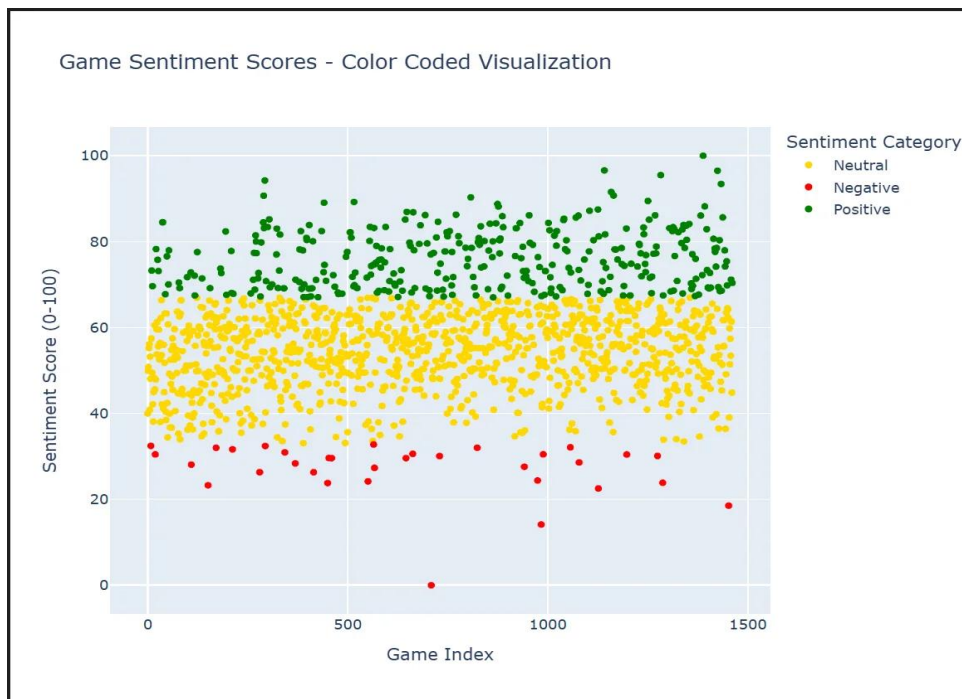


Figure 15. Game Sentiment Scores - Color Coded Visualization (0-100 Scale)

Figure 16 shows the top 30 games ranked by sentiment score with gradient colouring.

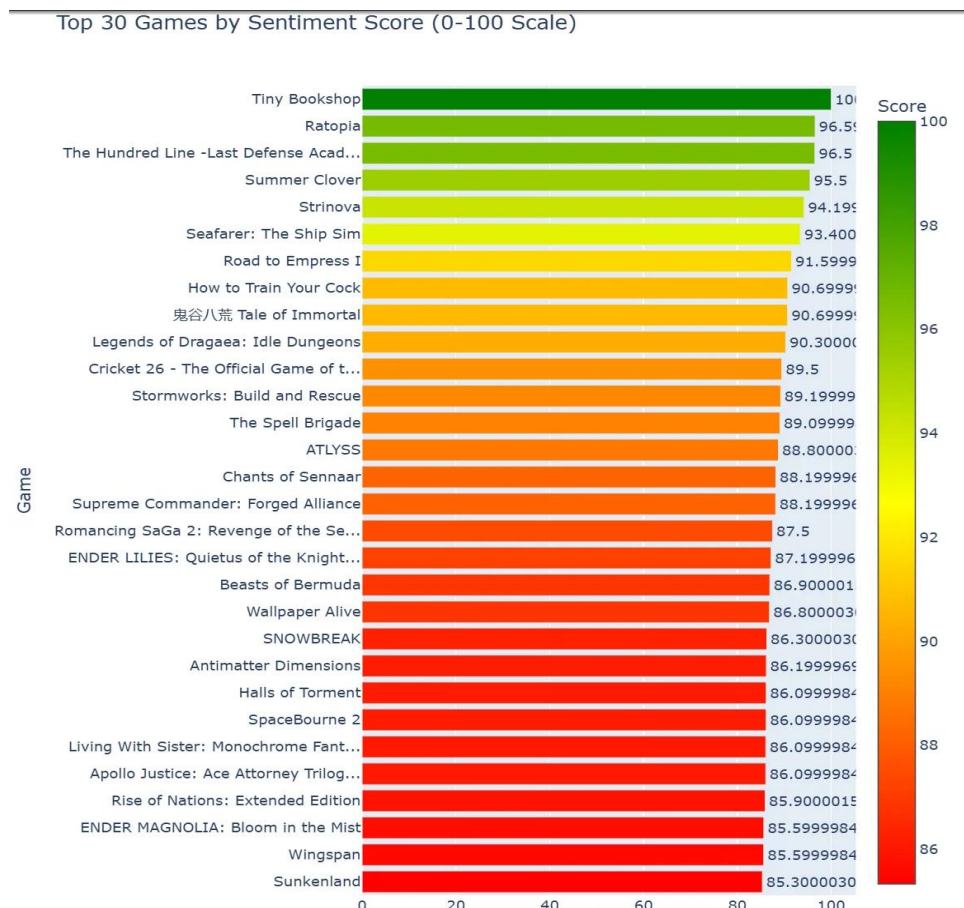


Figure 16. Top 30 Games by Sentiment Score (0-100 Scale)

Figure 17 shows the final sentiment scoring summary with category distribution.

```
=====
GAME SENTIMENT SCORING SUMMARY
=====

Total Games Analyzed: 1463

Sentiment Score Statistics (0-100 Scale):
  Minimum: 0.0
  Maximum: 100.0
  Mean: 58.5
  Median: 58.0
  Std Dev: 13.1

Sentiment Category Distribution:
  Negative (Red): 33 games (2.3%)
  Neutral (Yellow): 1085 games (74.2%)
  Positive (Green): 345 games (23.6%)

Color Coding:
  Red (0-33): Negative sentiment
  Yellow (34-66): Neutral sentiment
  Green (67-100): Positive sentiment

Top Game:
...
- results/predicted_vs_actual_100_scale.html
- results/platform_comparison.html

=====
Output is truncated. View as a scrollable element or open in a text editor. Adjust cell output settings...
```

Figure 17. Game Sentiment Scoring Summary (1,463 Games Analysed)

9. Troubleshooting

9.1 Common Issues

Error	Solution
ModuleNotFoundError: No module named 'tensorflow'	pip install tensorflow
FileNotFoundError: sentiment_modeling_dataset.csv	Run notebook 02 (feature engineering) first
praw.exceptions.ResponseException: 401	Check Reddit API credentials in config.py
MemoryError during training	Reduce batch size or close other applications

Table 4. Common Error Messages and Solutions