



# GameRx Research Notes

## GameRx: Your Digital Dose

This document captures the research foundations, data sources, and methodological decisions behind the GameRx project. It exists to make the work transparent, auditable, and explainable.

These notes document *why* specific tools, models, and frameworks were chosen, not just *what* was built.

---

## 1. Data Sources

### Steam Reviews Dataset

**Source:** Kaggle

**Dataset:** [AndrewMVD Steam Reviews](#)

#### What it contains

- Player-written review text
- AppID
- Timestamps and review metadata

This dataset is the emotional core of GameRx. It captures how players describe their experiences in their own words, without being constrained by predefined tags or ratings.

#### How it's used

- Emotion detection using NLP
- Review-level emotion scoring
- Emotion intensity and richness analysis
- Clustering emotional patterns
- Supporting genre–emotion analysis and visualizations

#### Processing notes

- Reviews are mapped to games using AppID and Steam metadata
  - Review text is cleaned, normalized, and filtered before emotion scoring
  - Emotion features are derived using the NRC Emotion Lexicon (NRCLex)
  - Review-level signals are later aggregated to the game level
- 

### Game Metadata Dataset

**Source:** Kaggle

**Dataset:** [Steam Games Dataset](#)

## What it contains

- Game title
- AppID
- Genre tags
- Release year
- Platform and developer information

This dataset provides the structural context needed to interpret emotional signals found in reviews.

## How it's used

- Genre normalization and parsing
- Feature engineering
- Game-level aggregation
- Linking reviews to genre and design context
- Interpreting emotional relief pathways

## Processing notes

- Game titles and AppIDs are aligned with the reviews dataset
  - Genres are cleaned, deduplicated, and expanded into structured lists
  - Primary genre and genre counts are derived for interpretation
  - Metadata supports clustering interpretation and fallback emotion labeling
- 

## 2. Emotion Detection Framework

### NRC Emotion Lexicon (NRCLex)

GameRx uses the **NRC Emotion Lexicon** via the NRCLex library to extract emotional signals from player-written review text.

#### Why NRC

- Captures discrete emotions (joy, sadness, anger, fear, trust, anticipation, etc.)
- Goes beyond binary sentiment
- Widely used in behavioral and psychological NLP research
- Interpretable and transparent

#### How it's applied

- Review text is tokenized and scored across NRC emotion categories
- Raw counts are normalized by review length
- Features created include:
  - Primary emotion
  - Emotion intensity
  - Emotion richness

#### Important constraint

NRC detects *expressed emotion*, not intent or outcome.

This means the model reflects **how players describe feeling**, not whether a game is objectively "good" or "bad."

This distinction is intentional and central to GameRx.

---

### 3. Emotion Theory & Framing

GameRx explicitly separates:

- **Emotion ≠ Sentiment**
- **Experience ≠ Evaluation**

A review expressing anger does not mean the game failed.

In many cases, anger, fear, or intensity are part of the *desired experience*.

This understanding informed every modeling choice in the project.

Rather than suppressing "negative" emotions, GameRx treats them as meaningful signals that can point to different kinds of emotional relief.

---

### 4. Hybrid Relief Model (Conceptual Basis)

To translate emotion signals into something actionable, GameRx introduces four **emotional relief pathways**:

- **Comfort** – soothing, safe, calming experiences
- **Catharsis** – release, venting, high-energy play
- **Distraction** – focus, immersion, attention shift
- **Validation** – emotional resonance, feeling seen

These categories are not psychological diagnoses.

They are **descriptive modes of how games help people cope**.

Relief pathways emerged through:

- Review emotion patterns
  - Genre context
  - Iterative auditing and correction
- 

### 5. Clustering & Modeling Choices

#### Clustering (Primary Modeling Method)

GameRx uses **KMeans clustering** to group emotional profiles.

**Why clustering**

- No labeled ground truth for “emotional game types”
- Allows patterns to emerge from the data
- Supports interpretability and exploration

## Implementation

- Emotion features derived from NRC scores
- Normalized inputs
- Multiple k values tested
- Final selection: **k = 5**

## Evaluation

- Elbow method
- Silhouette score
- Davies–Bouldin index
- Human interpretability of clusters

Clustering was chosen over classification because it revealed more meaningful emotional groupings without forcing labels prematurely.

---

## Classification (Exploratory Only)

Supervised models (e.g., Logistic Regression, Random Forest) were explored early in the project but intentionally dropped.

### Reason

- Required labels were subjective
- Results were less interpretable
- Did not align with the human-centered goals of the project

Clustering + hybrid relief modeling produced stronger, more honest insights.

---

## 6. Genre as Context, Not Ground Truth

Genres are used in GameRx as:

- Interpretive context
- Fallback signals when reviews are sparse
- Explanation tools for users

They do **not** directly drive recommendations.

A psychology-friendly genre mapping is used to help explain patterns (e.g., Strategy → Control, Simulation → Comfort), but emotional signals from reviews always take priority.

This avoids the assumption that:

“Genre = emotional outcome”

---

## 7. Safety, Filtering, and Ethics

GameRx includes explicit safety and content filtering.

Key principles:

- No games are deleted from the dataset
- Blocking is handled through flags, not removal
- A unified block registry ensures consistency
- Filtering logic is transparent and auditable

This design preserves analytical integrity while protecting users in the app.

---

## 8. Visuals as Validation

Visualizations are not decorative in GameRx.

They are used to **validate decisions**.

Visuals support:

- Emotion distribution checks
- Relief pathway balance
- Pre- and post-correction comparisons
- Cluster interpretability
- App behavior verification

Only visuals that reflect final, corrected logic are retained in the portfolio.

---

## 9. Limitations & Tradeoffs

This project acknowledges several constraints:

- Review language is subjective
- NRC does not capture nuance like sarcasm
- Low-review games require fallback logic
- Emotional states are simplified for usability

These tradeoffs are intentional and documented rather than hidden.