“Will This Game Be a Hit?”:

Explaining and Predicting Video Game Review Scores Using Machine Learning

Yuhan Wang, Kaiyue Deng

University of Southern California

DSC1-510: Principles of Programming for Data Science

Dr.Itay Hen

November 20, 2025

Project Introduction

Project Name: "Will This Game Be a Hit?": Explaining and Predicting Video Game Review Scores Using Machine Learning

Team Members: Yuhan Wang, Kaiyue Deng

Objective

This project aims to identify the key factors—such as genre, price, platform, publisher, and release timing—that correlate with high video game ratings. We will build a machine learning model to predict the potential success of new releases based on these attributes. The findings will provide developers with data-driven insights into market preferences and assist players in identifying promising titles efficiently.

Data Collection & Sources

We will collect data about video games from public web sources (e.g., Metacritic) using Python libraries requests and BeautifulSoup. For each game, we plan to collect: title, release date, platform(s), genre(s), price (if available), developer, publisher, overall rating/score, and reviews. For a subset of games, we will also collect aggregated user review information from the same source.

Analysis & Visualization

Our analytical approach consists of three phases:

Exploratory Data Analysis: Using Pandas and NumPy, we will examine correlations between structured features (e.g., price, publisher region) and game ratings. Visualizations, including boxplots and scatter plots (via Matplotlib) will highlight performance differences across genres and platforms.

Natural Language Processing (NLP): We will process review texts to extract keywords related to specific game aspects (e.g., story, gameplay mechanics, performance). We aim to quantify the sentiment of these aspects to determine which factors weigh more heavily for professional critics versus general users.

Predictive Modeling: Finally, we will combine structured metadata and NLP-derived features to train a Linear Regression model. This model will quantify feature importance and predict game ratings, evaluating how well these variables explain market success.