Content-Based Game Recommendation System

Leveraging Machine Learning to Deliver Personalized Gaming Experiences

Presented by: Maulik (Roll No. 12301052)Punjabi University, Patiala

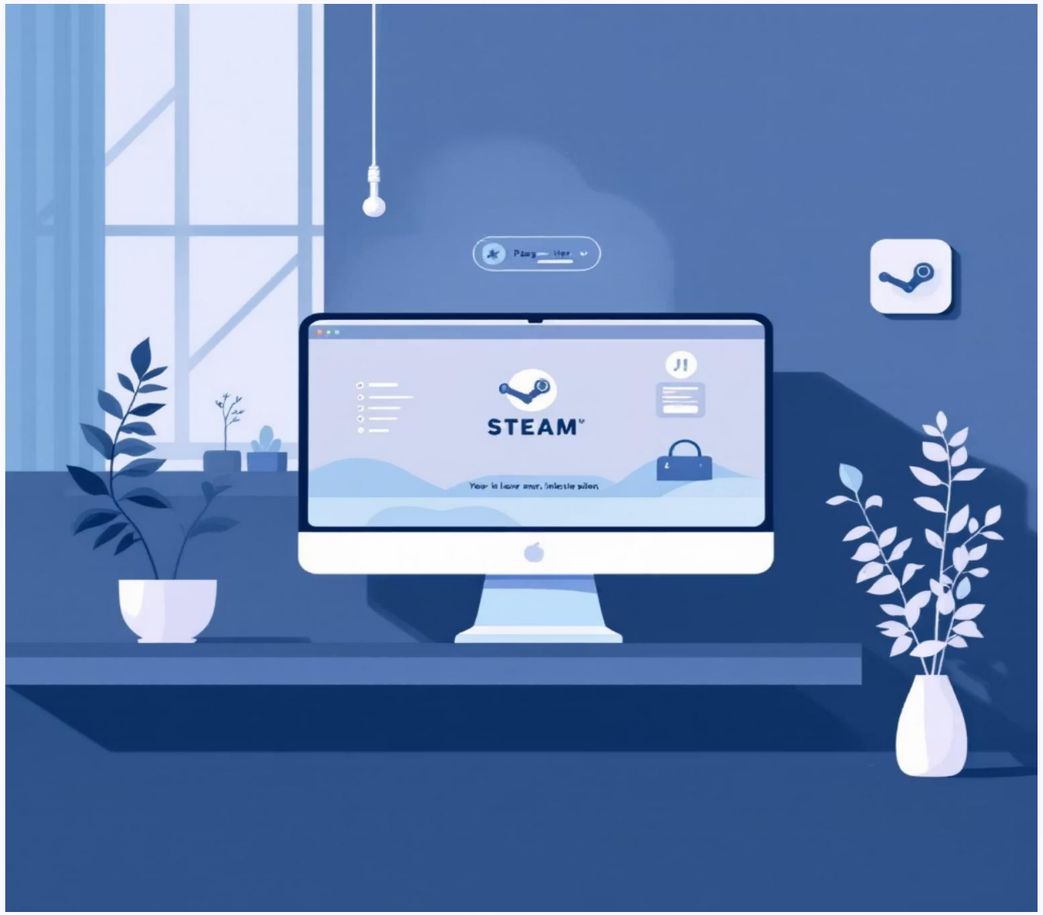
Project Introduction: Intelligent Game Discovery

What is Content-Based Recommendation?

A content-based recommender system analyzes the intrinsic characteristics and features of items to suggest similar content. Unlike collaborative filtering, which relies on user behavior patterns, our system examines game attributes–genres, tags, descriptions, and mechanics–to identify games with similar DNA.

The Gaming Discovery Challenge

With thousands of games available on Steam, players often struggle to find titles matching their preferences. Our system solves this by understanding what makes games similar and surfacing relevant recommendations based on content analysis rather than just popularity metrics.



System Capabilities

- Analyzes game metadata and descriptions
- Identifies content patterns and similarities
- Enables text-based game search
- Delivers personalized recommendations

Dataset Architecture & Integration

games.csv

Core dataset containing structured game information including titles, release dates, ratings, and pricing data. Serves as the primary data source with essential game identifiers.

games_metadata.json

Rich metadata repository with detailed descriptions, genre classifications, user-defined tags, and gameplay mechanics. Provides the textual content for similarity analysis.

Unified Dataset

Merged using **app_id** as the common key, creating a comprehensive game profile that combines structured data with rich textual features for analysis.

Sample Merged Data Structure

app_id	title	combined_features
570	Dota 2	MOBA Strategy Multiplayer Free-to-Play Competitive
730	Counter-Strike: GO	FPS Tactical Shooter Multiplayer Competitive Esports

Machine Learning Methodology: Unsupervised Content Analysis

Core ML Approach

Our system employs an **unsupervised learning** framework, requiring no labeled training data or user ratings. This approach analyzes inherent game characteristics to discover natural similarity patterns.

TF-IDF Vectorization

Term Frequency-Inverse Document Frequency transforms textual game features into numerical vectors. It identifies important words by weighing how frequently terms appear in a game's description against their rarity across all games, creating unique content signatures.

Cosine Similarity

Measures the angular distance between game vectors in high-dimensional space. Values range from 0 (completely different) to 1 (identical). Games with similar content patterns produce vectors pointing in similar directions, yielding high similarity scores.

Technical Advantage: This unsupervised approach eliminates cold-start problems and works effectively even for new games with no user interaction history.

System Architecture & Processing Pipeline

01

Data Ingestion

Load and merge games.csv with games_metadata.json using app_id as the join key

ĆČ

Text Preprocessing

Clean descriptions, remove stopwords, normalize text, and handle missing values

ĆĊ

Feature Engineering

Combine genres, tags, descriptions into unified text representations

04

TF-IDF Vectorization

Convert combined text into numerical feature vectors capturing content semantics

ĆD

Similarity Computation

Calculate pairwise cosine similarity scores between all game vectors

ĆĎ

Recommendation Generation

Retrieve and rank top-N most similar games based on similarity scores

The pipeline processes thousands of games in seconds, creating a responsive recommendation engine that updates as new games are added to the dataset. Each stage is optimized for performance while maintaining recommendation quality.

System Features & Capabilities

🎮 Game-Based Similarity Engine

Select any game from the Steam library and instantly receive recommendations for similar titles. The system analyzes content patterns to suggest games with comparable mechanics, themes, and gameplay styles.

🔍 Intelligent Text Search

Enter natural language queries describing your gaming preferences–genres, themes, or gameplay features. The system vectorizes your query and matches it against the entire game database to surface relevant recommendations.

🏷️ Dynamic Tag Extraction

Automatically identifies and displays the most relevant tags and categories for any game. Helps users quickly understand a game's core characteristics and discover related content through tag-based exploration.

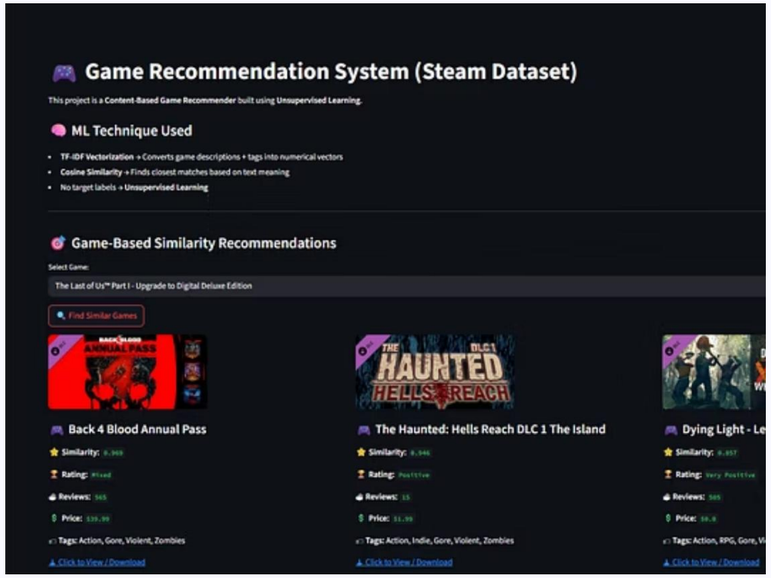
🖼️ Rich Visual Game Cards

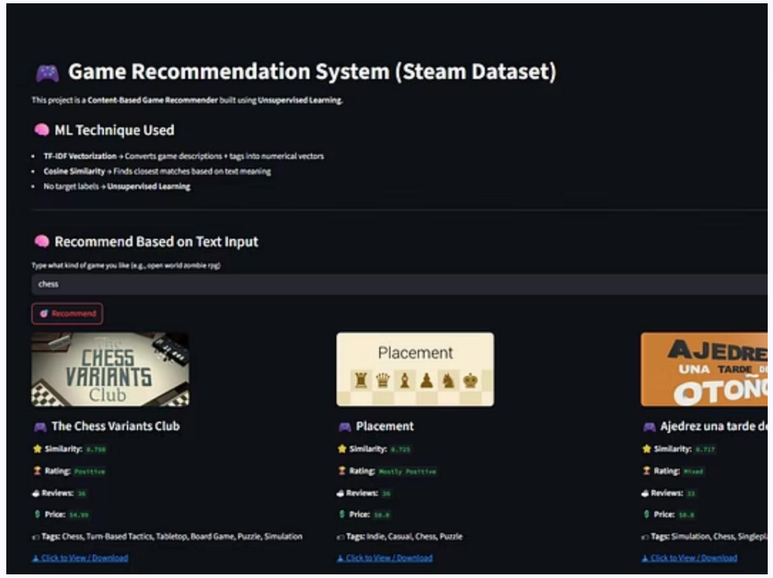
Each recommendation displays as an attractive card featuring the game's cover art, title, rating, release date, and key tags. Provides an engaging, Pinterest-style browsing experience.

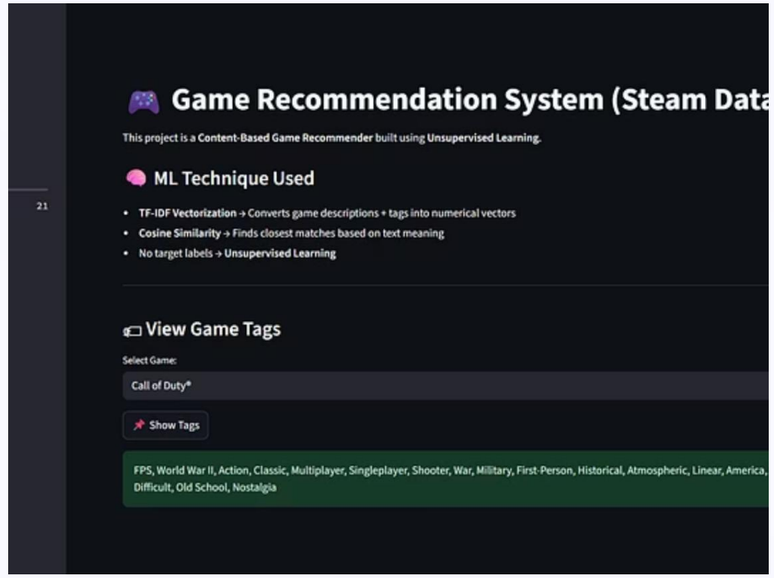
⚡ Streamlit Interactive UI

Built on Streamlit for rapid prototyping and deployment. Offers a clean, responsive interface with real-time recommendations, search suggestions, and seamless navigation across features.

System Interface & User Experience







Similar Games Discovery

The primary interface allows users to select a game and view similar recommendations displayed as visual cards. Each card shows cover art, metadata, and similarity scores, making it easy to discover new games aligned with preferences.

Text-Based Search

Natural language search enables users to describe what they're looking for–"multiplayer strategy games" or "story-driven RPGs"–and receive relevant matches ranked by content similarity to the query.

The Streamlit interface provides an intuitive, responsive experience across all features. Users can seamlessly switch between game-based recommendations, text search, and tag exploration. The visual design prioritizes clarity and engagement, making game discovery both effective and enjoyable.

Exploratory Data Analysis Insights

40K+

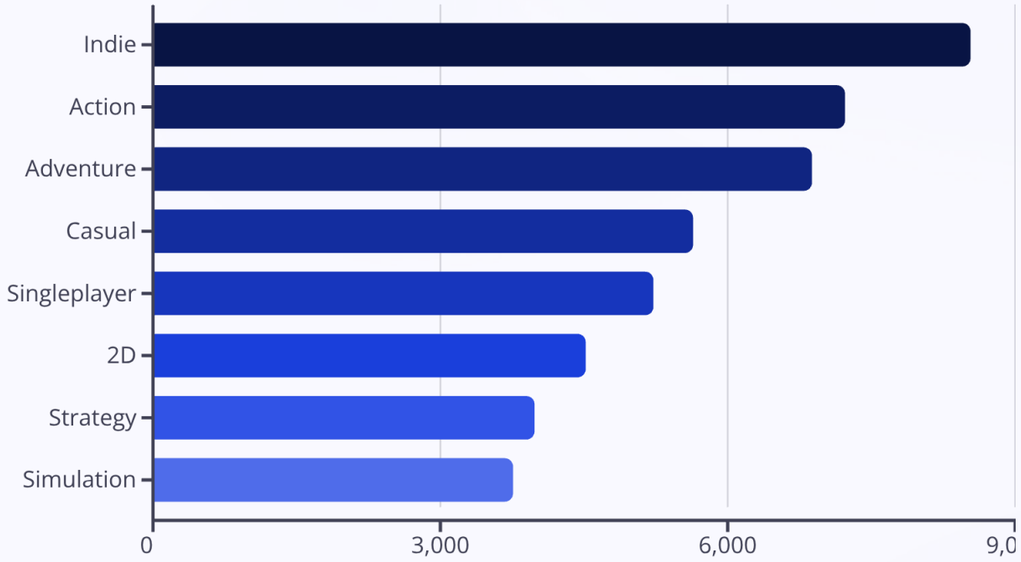
Total Games

Comprehensive Steam dataset spanning multiple genres and years

300+

Unique Tags

Diverse categorization enabling granular similarity matching



Top 20 Most Frequent Tags

Indie, Action, Adventure, Casual, Strategy, Simulation, RPG, Singleplayer, Multiplayer, 2D, Atmospheric, Story Rich, Great Soundtrack, Puzzle, First-Person, Pixel Graphics, Horror, Open World, Fantasy, Sci-fi

Rating Distribution

Most games cluster between **70-85%** positive ratings, indicating generally favorable reception across the dataset.

Genre Diversity

Strong representation across **Action, RPG, Strategy,** and **Indie** categories, ensuring varied recommendations.

Release Timeline

Dataset spans **multiple decades**, from classic titles to recent releases, providing comprehensive coverage.

Conclusions & System Performance

✅ High Similarity Accuracy

The TF-IDF and cosine similarity approach consistently identifies games with genuinely similar content profiles. Manual evaluation confirms that recommendations align well with genre, mechanics, and thematic elements.

⚡ Efficient Performance

The vectorization and similarity computation execute rapidly, processing thousands of games in under a second. The system scales effectively and maintains responsiveness even with large datasets.

🎯 No Labels Required

As an unsupervised learning system, it eliminates the need for extensive labeled training data or user rating histories. This makes it particularly valuable for new games and cold-start scenarios.

Key Achievements

- Successfully merged and processed 40,000+ game records
- Implemented robust text preprocessing and feature engineering
- Created an intuitive, visually appealing user interface
- Delivered real-time recommendations with high relevance

Future Enhancements

Potential improvements include incorporating user ratings for hybrid recommendations, adding collaborative filtering, implementing deep learning embeddings, and expanding to real-time Steam API integration.

Thank You

Questions & Discussion

Content-Based Game Recommendation System | Machine Learning Mini-Project

Maulik (Roll No. 12301052) Punjabi University, Patiala

This project demonstrates the power of unsupervised machine learning in creating intelligent recommendation systems. By leveraging TF-IDF vectorization and cosine similarity, we've built an effective tool for game discovery that requires no training labels and adapts naturally to new content.