GameSphere: A Smart Gaming Analytics & Recommendation Application

Devdatta Gandole, Nakshatra Sachinbhai Desai, Rajarajeswari Premanand, Tejashree Kishor Badve, Vineeth Rayadurgam
San Jose State University
Course: Data 226 - Big Data Technologies and Applications

***Abstract*—GameSphere is a web-based smart gaming analytics and recommendation system. It leverages PySpark and Hadoop's HDFS to deliver personalized game recommendations and insightful analytics based on sentiment analysis and collaborative filtering from a large dataset of Steam reviews.**

I. Introduction

GameSphere addresses the gaming community's need for personalized recommendations and insightful analytics. It analyzes approximately 8GB of Steam reviews to provide customized suggestions and graphical insights via a scalable backend using PySpark and HDFS.

II. Project Overview

GameSphere consists of two major components:

- **Gaming Analytics Dashboard**: Offers visualizations related to game popularity, review sentiment, and additional analytics.
- **Customized Game Suggestions**: Employs collaborative filtering using gaming tags and user activity to recommend games.

III. Methodology

A. Data Management

Data storage and processing involved:

- Data ingestion and preprocessing in HDFS.
- Structured storage setup as raw, processed, mappings, models, and outputs directories.
- Conversion of cleaned data into Parquet format for efficient access.

B. Data Cleaning and Preprocessing

The preprocessing steps included:

- Column standardization and missing value handling.
- Text normalization (lowercasing, removal of special characters).
- Data type conversions for accuracy and consistency.

C. Sentiment Analysis

Sentiment prediction utilized:

- Tokenization, TF-IDF vectorization.
- Logistic Regression model trained on English-language reviews.

D. Recommendation Engine

Collaborative filtering involved:

- ALS (Alternating Least Squares) algorithm.
- Conversion of user-game interaction data into numeric indices.
- Filtering unpopular games and outlier playtimes for quality recommendations.

IV. Technologies Used

- **Hadoop HDFS** for scalable data storage.
- **PySpark** for distributed processing.
- **Spark MLlib** for machine learning tasks.
- **Streamlit** for the frontend dashboard.
- **Pandas, Matplotlib, Seaborn** for exploratory data analysis and visualization.

V. Impact and Future Work

GameSphere significantly enhances user gaming experiences by providing personalized game recommendations and insightful analytics. It serves both gamers and developers by offering tools to analyze trends and user sentiments. Future extensions could include real-time recommendation updates and deeper integration with user social interactions.

VI. Conclusion

GameSphere successfully integrates big data analytics, machine learning, and web technologies to deliver an innovative platform for gaming insights and personalized recommendations, demonstrating practical use of modern big data technologies.

References

[1] Steam Reviews 2021 Dataset. Kaggle. https://www.kaggle.com
[2] Spark MLlib Documentation. Apache Spark. https://spark.apache.org/docs/latest/ml-guide.html