Project Title: ESRB Rating Prediction for Video Games

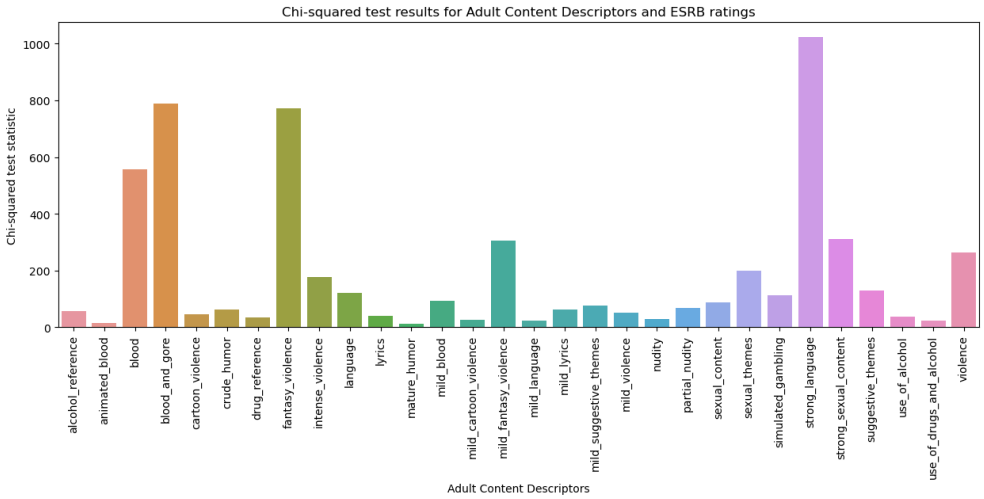
Project Overview

This project focuses on accurately predicting ESRB ratings for video games based on content descriptors and platform features. Utilizing machine learning models, the aim is to assist game developers, publishers, and ESRB in automating the rating process, ensuring compliance, and optimizing game targeting for specific age groups.

Context

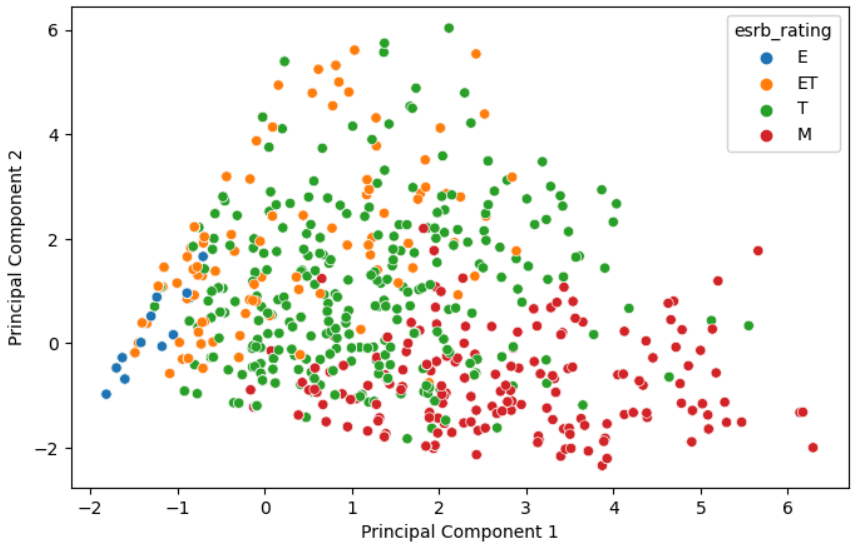
The Entertainment Software Rating Board (ESRB) assigns age and content ratings to video games in the United States and Canada. The project addresses the challenge of predicting ESRB ratings by developing a machine learning model trained on a dataset containing 34 content descriptors and platform details for 1895 games.

Exploratory Data Analysis (EDA)



Features like **strong_language, blood_and_gore, blood and fantasy_violence** are among the most strongly associated features.

Principal Component Analysis (PCA)

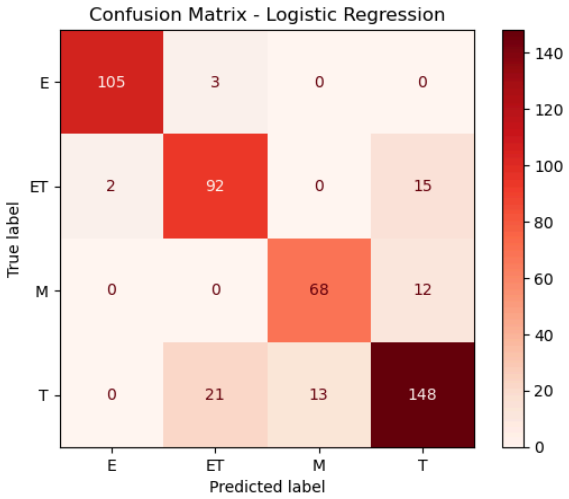


PCA revealed distinct clusters for 'E' and 'M' ratings, while 'ET' and 'T' clusters appeared to overlap. Insights from PCA guide the model development phase.

Modeling

Three models were implemented and evaluated:

1. **Multinomial Logistic Regression:**
   - Train Accuracy: 0.8653
   - Test Accuracy: 0.8601
2. **Gradient Boosting:**
   - Train Accuracy: 0.9211
   - Test Accuracy: 0.8497
3. **Random Forest:**
   - Train Accuracy: 0.9159
   - Test Accuracy: 0.8559



Conclusion

Multinomial Logistic Regression emerged as the preferred model with an accuracy of **0.86**, showcasing stability between training and test datasets. The model's capability to predict ESRB ratings, along with its efficiency, makes it a valuable tool for game developers, publishers, and the ESRB in optimizing game content and automating the rating process.

Recommendations

1. Implement the Multinomial Logistic Regression model for accurate ESRB rating predictions.

2. Utilize model insights, such as the strong association of features like 'strong_language,' 'blood_and_gore,' 'blood,' and 'fantasy_violence,' to tailor game content for specific age groups and align with ESRB guidelines.
3. Explore the potential of automating the ESRB rating process with the developed model.

Future Work

Future enhancements could involve incorporating additional features or experimenting with advanced modeling techniques to further refine predictive capabilities. Continuous monitoring and updates to the model will ensure its adaptability to evolving gaming content and industry standards.