ESRB Rating Prediction for Video Games

Agenda

- ESRB Rating Prediction for Video Games
- Context
- Exploratory Data Analysis (EDA)
    - Principal Component Analysis (PCA)
- Modeling
    - Multinomial Logistic Regression
    - Gradient Boosting
    - Random Forest
- Conclusion
- Recommendations
- Future Work

ESRB Rating Prediction for Video Games

This project aims to predict ESRB ratings for video games using machine learning models based on content descriptors and platform features.

The goal is to automate the rating process, ensure compliance, and optimize game targeting for specific age groups.

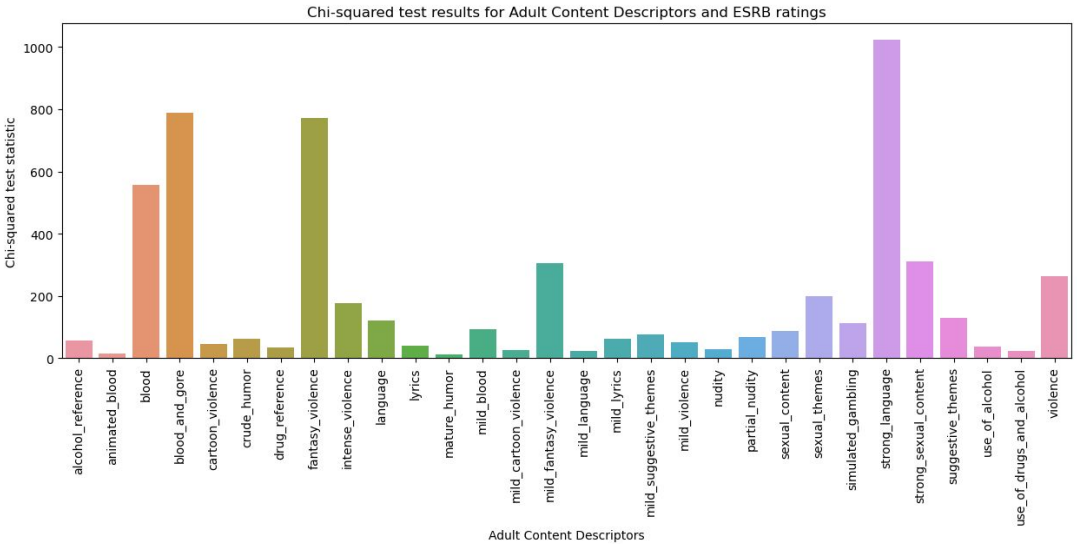
The prediction model, once implemented, will be a valuable tool for game developers, publishers, and the ESRB.

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 for video games based on content descriptors and platform features.

Exploratory Data Analysis (EDA)

Association of Features with ESRB Ratings



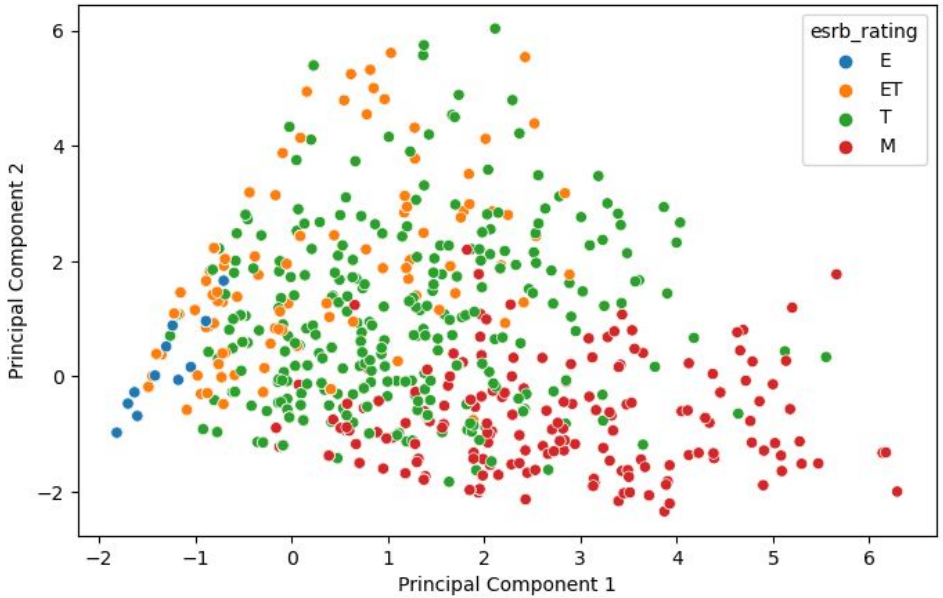
Key Findings from EDA:

- The EDA revealed that features like **'strong_language**,' **'blood_and_gore**,' and **'fantasy_violence'** are among the most strongly associated with certain ESRB ratings.
- Principal Component Analysis (PCA) showed distinct clusters for 'E' and 'M' ratings, while 'ET' and 'T' clusters appeared to overlap.
- These insights guide the model development phase, influencing the selection of appropriate features and algorithms.

Principal Component Analysis (PCA)

Variance Explained by Principal Components

PCA revealed distinct clusters for 'E' and 'M' ratings, while 'ET' and 'T' clusters appeared to overlap, indicating potential classification challenges in these areas.



Modeling

1

Three machine learning models were implemented and evaluated: **Multinomial Logistic Regression, Gradient Boosting, and Random Forest.**

2

The models were assessed based on their train and test accuracies, with Multinomial Logistic Regression achieving the highest overall performance.

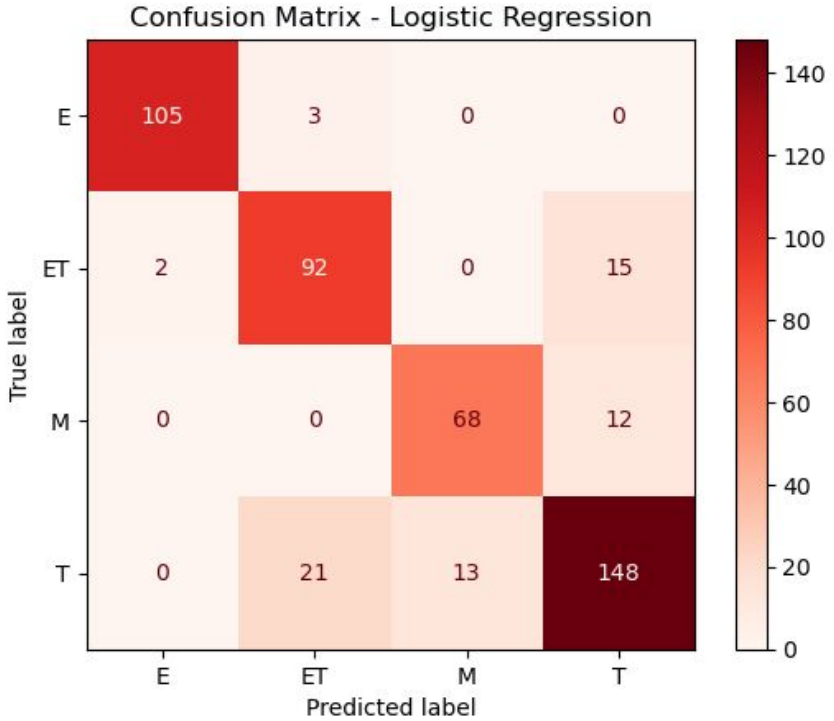
3

The **Multinomial Logistic Regression** model showcased the preferred balance between accuracy and stability, making it the optimal choice for ESRB rating prediction.

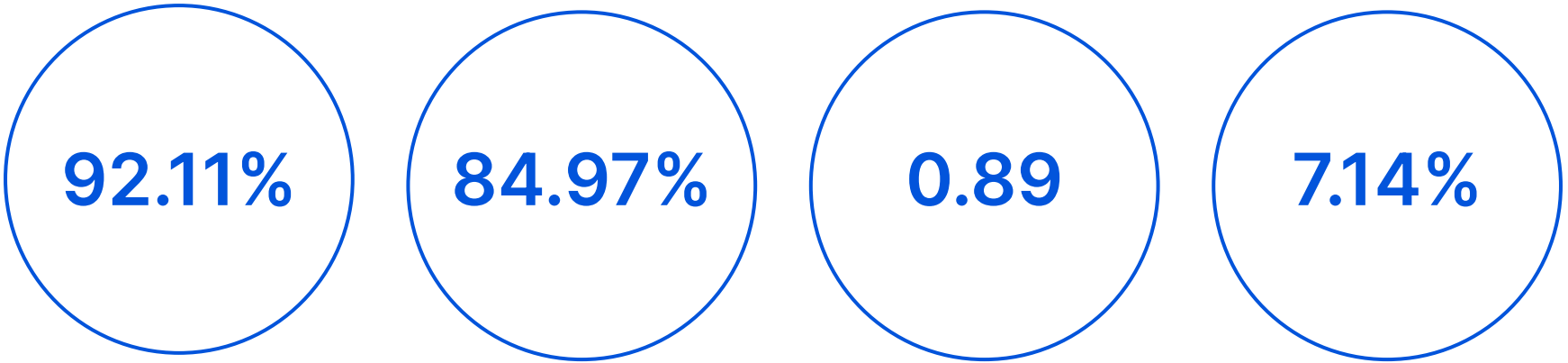
Modeling - Multinomial Logistic Regression

86.53%	86.01%	0.92	0.32
Train Accuracy	Test Accuracy	AUC Score	Cross-Entropy Loss

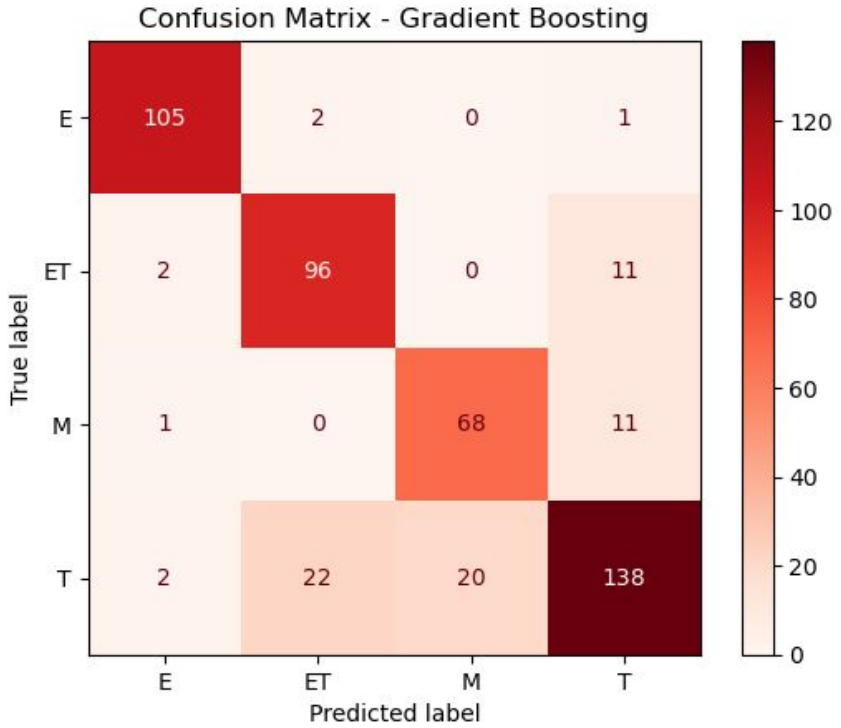
Modeling - Multinomial Logistic Regression



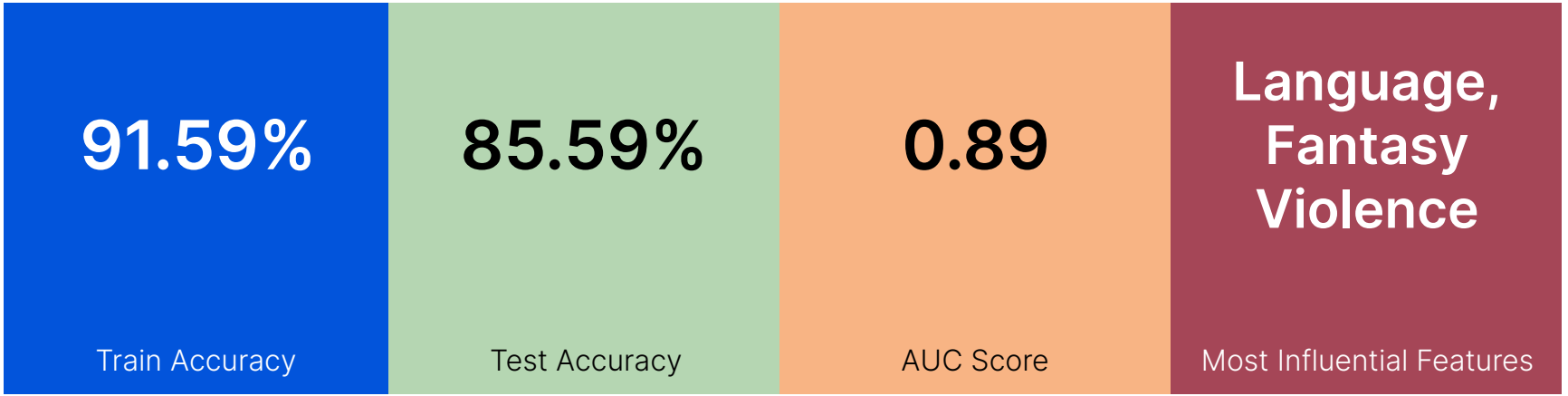
Modeling - Gradient Boosting



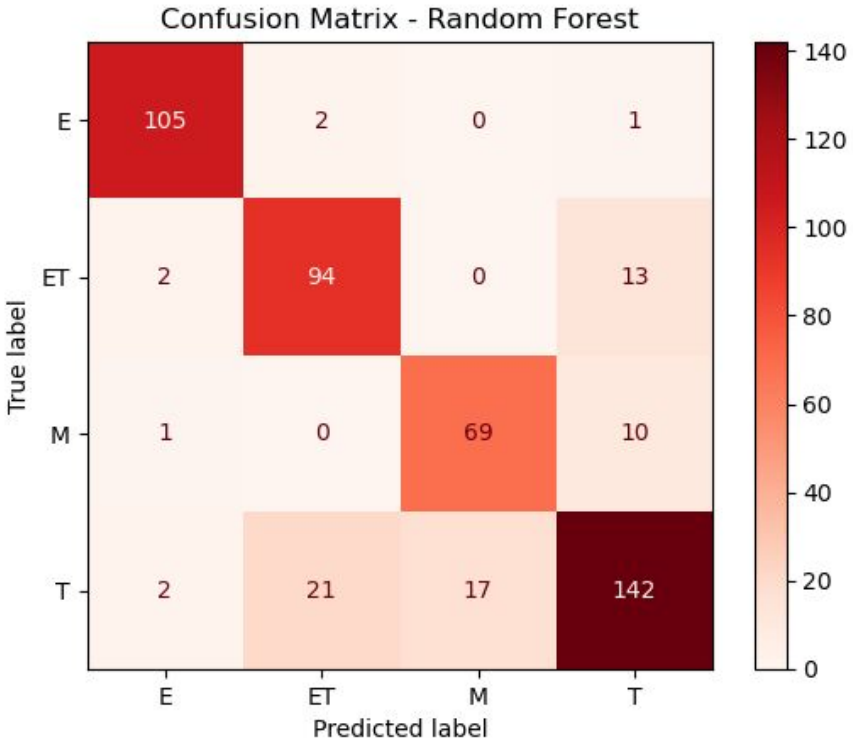
Modeling - Gradient Boosting



Modeling - Random Forest



Modeling - Random Forest



Conclusion

- The Multinomial Logistic Regression model, with an accuracy of **0.86**, demonstrates stability between training and test datasets.
- The model shows capability in predicting ESRB ratings, showcasing efficiency and reliability.
- The automation of the rating process, if implemented, could bring substantial benefits to game developers, publishers, and the ESRB.
- The model presents a valuable tool for the industry in optimizing game content and targeting specific age groups.

Recommendations

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

Strong association of features like **'strong_language**,' **'blood_and_gore**,' **'blood**,' and **'fantasy_violence**,' can be used to tailor game content for specific age groups and align with ESRB guidelines.

Automate the ESRB rating process with the developed model to optimize efficiency and compliance.

Continuously monitor and update the model to ensure its adaptability to evolving gaming content and industry standards.

Future Work

Additional Feature Incorporation

- Consider incorporating new content or platform features that may impact ESRB ratings.
- Explore data sources to include additional content descriptors for a more comprehensive model.
- Assess the impact of adding new features on the model's performance and accuracy.

Advanced Modeling Techniques

- Experiment with more complex and specialized machine learning models.
- Investigate the use of natural language processing for text-based content descriptors.
- Evaluate the potential of deep learning for ESRB rating predictions.