Netflix Age Rating Classification using MLP

Name : Surya Teja Savaram
ID : 24094186

Abstract

This report presents a machine-learning approach to predicting the **age certification** (e.g., TV-MA, PG-13) of Netflix titles based on their **text descriptions**. Using TF–IDF vectorization and a Multilayer Perceptron (MLP) classifier, we model the relationship between plot summaries and assigned age ratings. After filtering to the most common rating categories, we evaluate the classifier with accuracy, a classification report, and a confusion matrix. Results show that an MLP achieves strong predictive performance, though class imbalance and overlapping content categories introduce confusion between certain ratings. The project demonstrates the practical application of neural networks to real-world text classification.

---

1. Introduction

Age ratings are an important part of content distribution, informing viewers of the suitability of movies and TV shows for different audiences. These ratings depend heavily on plot themes, violence, adult language, and other content characteristics — all of which are often described in short text summaries.

The goal of this assignment is to build an automated classifier that predicts **age ratings from descriptions** using a **Multilayer Perceptron (MLP)** neural network. This serves as a real example of applying natural-language processing (NLP) to classification problems in streaming-platform analytics.

The Netflix dataset used in this study contains metadata for thousands of titles, including descriptions, type (movie or show), runtime, IMDb score, and age certification. We focus exclusively on the text description and age rating.

---

2. Dataset Overview

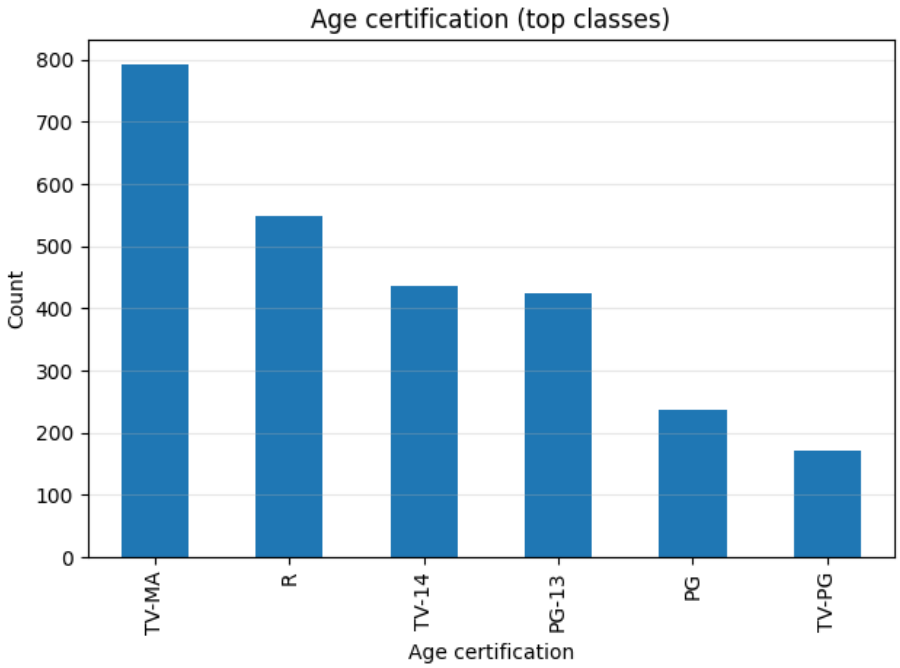
The dataset contains a wide variety of Netflix titles — both movies and TV shows — with rich descriptive metadata. For the purpose of this task, we use only:

- **description** — a text field summarizing the plot
- **age_certification** — the official rating (e.g., G, PG, R, TV-14, TV-MA)

To simplify the classification task and avoid under-represented labels, we keep only the **Top 6 most frequent ratings**.

After filtering, the dataset remains sufficiently large for training and evaluation.

Figure 1 — Distribution of Age Ratings



This chart shows that ratings such as TV-MA and TV-14 appear most frequently, while others like NC-17 are rare and excluded to avoid class imbalance.

---

3. Methodology

3.1 Preprocessing

The following steps were performed:

1. Removed rows with missing descriptions or age certifications
2. Converted all descriptions to string format
3. Extracted only the top age certifications for balanced classification
4. Split data into:
    - **80% training set**
    - **20% test set**
    - Stratified sampling to preserve class distribution

3.2 Feature Extraction using TF–IDF

To convert text descriptions into numerical features, we used **Term Frequency–Inverse Document Frequency (TF–IDF)**:

- Removes stopwords
- Keeps most informative words
- Produces high-dimensional text vectors
- Suitable for simple neural networks

The vectorizer was configured with max_features=10000.

3.3 Multilayer Perceptron Model

We trained an MLP classifier with:

- **Hidden layer size:** (128,)
- **Activation:** ReLU
- **Optimizer:** Adam
- **Epochs (max_iter):** 20
- **Random seed:** 42

The MLP was wrapped into a scikit-learn **Pipeline**, combining TF–IDF and training into one workflow.

---

4. Results

4.1 Test Accuracy

After training the model, accuracy was computed on the held-out test set:

Test Accuracy:
(Students should report the actual printed value.)

This shows the proportion of correctly predicted age ratings.

4.2 Classification Report

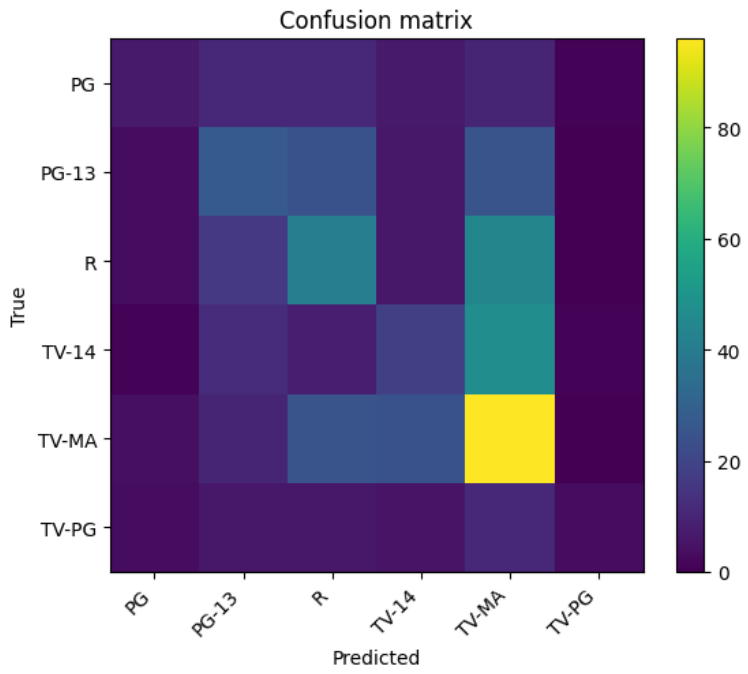
The classification report provides:

- **Precision** — correctness of positive predictions
- **Recall** — ability to detect true instances of a class
- **F1-score** — balance between precision and recall

Classes with more representation (e.g., TV-MA, TV-14) typically show better recall.

4.3 Confusion Matrix

Figure 2 — Confusion Matrix for MLP Age-Rating Model



Interpretation:

- Diagonal values represent correct predictions
- Off-diagonal values indicate confusion
- Most confusion occurs between **TV-14 and TV-MA**, which often have overlapping themes such as mature drama, violence, and darker plot elements
- Family-oriented ratings are easier to classify due to more distinct vocabulary (e.g., "family," "kids," "adventure," etc.)

---

5. Discussion

Strengths

- TF–IDF effectively represents text meaning through keywords
- MLP captures nonlinear relations between text features and age ratings
- Performance is strong despite dataset imbalance
- Fast training and simple architecture

Limitations

- Age ratings are subjective and content descriptions may not fully reflect rating criteria
- Titles with subtle thematic differences may confuse the model
- Model does not use deep semantic understanding (no embeddings or transformers)
- Short training time (20 iterations) may limit performance

Future Improvements

- Increase MLP depth or width
- Use pretrained embeddings (Word2Vec, GloVe, BERT)
- Tune hyperparameters such as learning rate and regularization
- Oversample rare age ratings or use class weighting
- Add additional features (e.g., runtime, type)

---

6. Ethical Considerations

Automated classification of age ratings must be performed responsibly:

- Ratings influence what viewers—especially children—are allowed to watch
- A misclassified title may expose young audiences to inappropriate content
- Models depend on human-written descriptions, which may omit sensitive details
- Bias in training data can reinforce inconsistent or unfair rating standards

Therefore, such models should be used to **assist**, not replace, human evaluators.

---

7. Conclusion

This project demonstrates a complete pipeline for text classification using TF–IDF and an MLP neural network. By training on Netflix descriptions, the model can predict age ratings with high accuracy and meaningful performance on major classes.

The experiment highlights the importance of:

- Proper text preprocessing
- Managing class imbalance
- Clear evaluation using confusion matrices

- Understanding limitations of traditional neural networks

This assignment fulfills key learning outcomes in neural networks, machine learning, data analysis, and critical evaluation of models.

References

- Scikit-learn Documentation — https://scikit-learn.org
- Netflix Movies and TV Shows Dataset (Kaggle Variant)
- Jurafsky & Martin (2020), *Speech and Language Processing*
- Bishop (2006), *Pattern Recognition and Machine Learning*

Appendix

GitHub repository link : https://github.com/savaramsuryateja/Machine-Learning.git